AD_____

Award Number: W81XWH-08-1-0636

TITLE: PREMALIGNANT GENETIC AND EPIGENETIC ALTERATIONS IN TUBAL EPITHELIUM FROM WOMEN WITH BRCA1 MUTATIONS

PRINCIPAL INVESTIGATOR: ANTON KRUMM

CONTRACTING ORGANIZATION: University of Washington
Seattle WA 98195

REPORT DATE: October 2011

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-10-2011	Final	15 SEP 2008 - 14 SEP 2011

4. TITLE AND SUBTITLE
PREMALIGNANT GENETIC AND EPIGENETIC ALTERATIONS IN TUBAL EPITHELIUM FROM WOMEN WITH BRCA1 MUTATIONS

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-08-1-0636

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
ANTON KRUMM

E-Mail: akrumm@u.washington.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Washington
Seattle WA 98195

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Increasing evidence suggests that many types of ovarian cancers originate within thefallopian tube. The scope of this Translational Partnership project is to define a uniquepremalignant gene expression profile and to identify causal epigenetic relationships. Ouranalyses have identified a premalignant expression signature that reflects early steps inovarian carcinogenesis. While genes differentially expressed in BRCA1 normal Fallopian Tubeepithelia and BRCA1 ovarian carcinoma were investigated in the Swisher lab, we have furtherrefined evidence for antagonistic action of DNA methylation and binding of the nuclear factorCTCF at genomic loci. Using a quantitative DNA methylation assay (EpiTYPER™) and lasercapture microdissection, we profiled several gene loci that are part of the premalignantsignature in ovarian cancer cell lines. While the majority of loci did not reveal anydifference between BRCA1 cancers and normal risk fallopian tube epithelia, a small fractionof cancers indicate increased methylation of the CDKN1C tumor suppressor gene. Contrary topublished reports, our data exclude a role for LOH or estrogen-inducible antisense RNAs.

15. SUBJECT TERMS
None provided.

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	70	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

Although it has been proposed that ovarian cancer originates from the surface epithelium of the ovary and/or the epithelial lining of ovarian inclusion cysts, there have been few reports of pre-neoplastic or early lesions at these sites. Instead, there has been increasing evidence that many ovarian cancers originate within the fallopian tube. Although the lifetime risk of ovarian cancer in the general population is 1-2%, women who inherit a mutation in the BRCA1 gene have up to a 50% lifetime risk of ovarian cancer. These high-risk women are frequently discovered to have occult neoplasms at the time of risk-reducing salpingo-oophorectomy, and 57-100% of these lesions are discovered in the fallopian tube. Our tissue bank included frozen fallopian tube tissue from women with BRCA1 mutations found to have occult fallopian tube carcinomas on final pathologic examination. We hypothesized that the histologically normal fallopian tube epithelium from these women would possess a unique gene expression profile which would reflect early disruptions in gene expression contributing to the development of carcinoma. We proposed to investigate the novel idea that altered expression of some candidate genes was due to changes in DNA methylation status at target motifs for the zinc-finger protein CTCF.

BODY

In the first year, we completed the work outlined in our statement of work with minor modifications. The objective of the first year was to complete laser capture microdissection, RNA preparation and amplification and expression array analyses for BRCA1 ovarian cancers, fallopian tube epithelium (FTE) from women at normal risk of ovarian cancer and FT from women with BRCA1 mutations, then obtain a list of candidate genes that may contribute to early ovarian carcinogenesis (Figure 1 and attached manuscript). We have completed that objective as detailed below. We originally planned to do 60 samples. However, due to the markedly decreased budget we were unable to complete analyses on that many samples and reduced our work to a total of 48 samples. This allowed us to achieve our objective while retaining adequate funds to carry out the epigenetic work outlined for Year 2.

To determine if changes in gene expression profiles within the histologically normal fallopian tube epithelium of BRCA1 mutation carriers would overlap with the expression profiles in BRCA1-mutated ovarian carcinomas and represent a BRCA1 preneoplastic signature, we performed laser capture microdissection of frozen sections to isolate neoplastic cells or histologically normal fallopian tube epithelium. Expression profiles were generated on Affymetrix U133 Plus 2.0 gene expression arrays. Normal-risk controls were 11 women with wild-type alleles of BRCA1 and BRCA2 (WT-FT). WT-FT were compared with histologically normal FTE from seven women with deleterious BRCA1 mutations who had foci of at least intraepithelial neoplasm within their fallopian tube (B1-FTocc). WT-FT samples were also compared with 12 BRCA1 ovarian carcinomas (B1-CA). The comparison of WT-FT versus B1-FTocc resulted in 152 differentially expressed probe sets, and the comparison of WT-FT versus B1-CA resulted in 4079 differentially expressed probe sets. The BRCA1 preneoplastic signature was composed of the overlap between these two lists, which included 41 concordant probe sets. Genes in the BRCA1 preneoplastic signature included several known tumor suppressor genes such as CDKN1C and EFEMP1 and several thought to be important in invasion and metastasis such as E2F3. The expression of a subset of genes was validated with quantitative reverse transcription–polymerase chain reaction and

immunohistochemistry. The work for Year 1 was performed in the Swisher lab as planned. The results of these studies were published in *Neoplasia* at the end of 2010. That manuscript is attached to this report.

Changes in the methylation status of DNA have the potential to serve as an early detection marker for malignancies. Previous work in several labs including ours has revealed that epigenetic aberrations including methylation of CTCF target sites are a key event in carcinogenesis [[1], reviewed in [2]]. CpG methylation of CTCF target motifs inhibits binding of CTCF and permits spreading of DNA methylation and subsequent silencing of genes such as tumor suppressor genes [3]. In the second and third (extension) years we completed the work outlined in our statement of work with minor modifications. To test the epigenetic contribution to the malignant transformation in the ovary, we pursued two objectives: (1), determine a role for DNA methylation and CTCF binding in the deregulation of specific genes including those identified by the Swisher lab and (2), evaluate and validate the contributions of alterations in DNA methylation and CTCF binding.

The epigenetic analyses were performed in the Krumm lab and the Swisher lab identified tissue samples and performed laser capture microdissection to obtain DNA samples on 40 normal FT and malignant samples. The Krumm laboratory requested a no cost extension to complete the epigenetic analyses. FTE and ovarian carcinomas were obtained from an IRB approved tissue bank. Laser capture microdissection (LCM) of formalin fixed, paraffin embedded sections was used to isolate neoplastic cells or histologically normal FTE. Total DNA was obtained using the Picopure DNA isolation kit (Arcturus).

Because the exact role for DNA methylation in controlling CTCF binding is poorly defined, we first examined the contribution of individual cytosine residues to binding of CTCF at several genomic loci. CTCF binding sequences derived from several genes including the human MYC oncogene as well as the IGF2 gene and the IGF2BP1 gene were tested for their ability to recruit CTCF in vitro using immobilized template assays (Figure 3). DNA templates either methylated or unmethylated were linked to magnetic beads, incubated with nuclear extract, washed, and tested for association with CTCF by western blotting. As shown in Figure 3B, templates containing wildtype CTCF binding sequences derived from the IGF2BP1 gene (IGF2BP1 wt, lower panel of Figure 3B) efficiently recruit CTCF. In contrast, CTCF binding was severely reduced when the target motifs were mutated by three base substitutions (IGF2BP1 mut, Figure 3B). To establish whether binding of CTCF to its target motifs is inhibited by cytosine methylation, we tested immobilized templates after SspI-mediated methylation of cytosine residues *in vitro*. We examined CTCF motifs containing different CpG content such as the myc site A [3] as well as the B1 sequence of the ICR of the human IGF2/H19 locus [4]. Cytosine methylation at the human B1 sequence is known to inhibit binding of CTCF [4]. Consistent with this, recruitment of CTCF to immobilized templates containing the B1 sequence or the myc site A is highly sensitive to DNA methylation. (Figure 3A, upper panel). In contrast, CpG methylation of the IGF2BP1 motif has no effect on CTCF recruitment. Replacement of the IGF2BP1 core motif by the CTCF-binding sites of the chicken FII insulator element yields similar results. However, CTCF binding becomes sensitive to CpG methylation upon modification of the core motif into human B1 sequence. In combination, our experiments indicate that the inhibition of CTCF binding is not only dependent on DNA methylation but is also dependent on additional features of the motif including the number and position of cytosine residues. These results were published in *Epigenetics and Chromatin* in August of 2011. This manuscript is also attached to this report.

To further investigate a role for cytosine methylation in the inhibition of CTCF binding and deregulation of gene expression, we tested several sites at the HOX gene locus *in vivo* in cancer cell lines established from ovarian, breast and prostate tissue (Figure 4). The HOXA gene cluster is a family of homeotic genes that encode transcription factors frequently inactivated in cancer cell types. The HOX gene domain contains several CTCF sites (hx1-hx5) previously identified in our ChIP-Chip analysis in the breast epithelial cell line HBL100. Importantly, CTCF binding at hx1 is absent in the prostate epithelial cell line PC3 [5]. Genomic sequencing of the hx1 region revealed complete sequence identity at this site in HBL100 and PC3 cells, suggesting that epigenetic mechanisms account for the loss of CTCF binding in PC3 prostate cancer cells. Using a combination of methylation-sensitive restriction enzymes and PCR we analyzed the level of DNA methylation in the ovarian cancer cell line A2780 and compared it to the prostate cancer cell lines PC3 and C4-2. These experiments identified DNA methylation at the hx1 binding site in A2780 and PC3 cells but not in HBL100 cells and C4-2 cells. Most importantly, these studies further confirm a correlation of DNA methylation and loss of CTCF binding; while hx1 in the prostate cell line C4-2 is both unmethylated and bound by CTCF, hx1 in the A2780 and PC3 cell lines is methylated and not bound by CTCF. These data further support our hypothesis that epigenetic mechanisms and loss of CTCF binding contribute to reprogramming of gene expression during disease progression.

To address the potential role of CTCF binding and DNA methylation in deregulation of those genes identified by the Swisher lab, we scanned the genomic regions harboring candidate genes for known CTCF binding sites. Importantly, the majority of candidate genes are associated with one or more CTCF sites in a sequence space of 100 kb surrounding candidate loci. For only six loci is the closest CTCF binding site located more than 100 kb away. The distribution of CTCF sites across the subset of premalignant signature genes is similar to the distribution found genome-wide: About one half of CTCF sites are located in intergenic regions, with an average distance of approximately 47 kb. About 20% CTCF sites are located at transcription start sites, and 34% are located within introns and exons. Three examples of loci with CTCF binding sites in the vicinity of the genes under study are shown in Figure 5.

To obtain initial data on differential binding of CTCF at premalignant signature genes, we performed methylation-sensitive PCR on genomic DNA from several cancer cell lines including the ovarian cancer cell line OVCAR3. This analysis takes advantage of the methylation-sensitive restriction enzyme AciI that digests only unmethylated genomic regions, eliminating templates for subsequent PCR. Thus, while unmethylated regions yield no PCR product, methylated regions are protected from restriction digest and produce amplified DNA fragments. Using this approach, we investigated the methylation status at PAK3, JAG1, and LOC388798 gene loci. As shown in Figure 6 and 7, the CTCF binding region in PAK3 is methylated in the ovarian cancer cell line OVCAR3 but is un-methylated in the prostate cell line LnCaP. In contrast, our analyses at the LOC388798 on chromosome 20 revealed that this region is unmethylated in all cell lines tested.

Our technical objective 2 originally included methylation analysis using methylated DNA immunoprecipitation (MeDIP) and microarrays tiling through gene loci differentially expressed between normal tubal epithelium and BRCA1 carcinomas. However, our previous experience indicated that this approach limited our ability to quantify methylation levels. Moreover, while other approaches including extensive bisulfite sequencing can quantitatively reveal differential methylation between normal and tumor cells, these methods do not permit DNA methylation analyses at a high-throughput level. To accommodate highly quantitative and efficient analyses of methylation at differentially expressed gene loci, we employed EpiTYPER, a quantitative DNA methylation analysis using the MassARRAY® system. This approach combines bisulfate-mediated

base-specific cleavage of methylated DNA and matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF) previously introduced for SNP discovery. This approach includes a PCR step in which bisulfite-treated genomic DNA is amplified with primers containing a T7 promoter sequences. After transcribing DNA by T7 polymerase, RNA is cleaved in a base-specific manner and analyzed by MALDI-TOF mass spectrometry. This approach generates quantitative results for each cleavage product with a standard deviation of 5%, an important feature and precision that is not available with other approaches such as MeDIP-Chip analyses. Moreover, the EpiTYPER platform is capable of detecting methylation levels as low as 5% in sample mixtures and is thus highly sensitive and useful for the precise characterization of epigenetic changes in cancer phenotypes.

A very important step in epigenetic analyses using the EpiTYPER technology is the selection of PCR primers and the establishment of robust PCR conditions. Treatment of DNA with sodium bisulfate results in the complete transformation of unmethylated cytosines to uracil. The chemically converted cytosines are amplified by PCR as thymines. Analysis of these PCR products reveals the initial methylation profile of the region of interest. A technical advantage of the method resides in the use of PCR, which allows for analysis of samples with very low DNA content. However, PCR amplification can often be the most difficult with the challenge residing in the specific amplification of bisulfite-treated DNA. High redundancy of the target sequences as reflected by the original G/C richness creates long stretches of thymines, which are often difficult for polymerases to faithfully replicate. Moreover, DNA fragmentation during bisulfate treatment leads to an empirical upper size limit of the PCR amplicon of 400–500 bp. Indeed, only short amplicons are amplified and the need for nested primers and a second round of PCR is often necessary.

To overcome these technical challenges, primer design is crucial since dimer formations are greatly facilitated by the T/A richness of the sense and antisense oligos, respectively. Moreover, primers designed for bisulfite-treated templates frequently generate non-specific PCR products because of mispriming in the highly redundant genome generated by bisulfate treatment. Although several primer-design algorithms exist for amplification of bisulfite-treated DNA the identification of reliable primer combinations remains difficult.

Primer design coupled with DNA degradation during tissue fixation/extraction and bisulfite treatment challenge efficient amplification. Thus, we spent a significant amount of effort both to establish experimental conditions that limit DNA degradation and to select primer pairs that efficiently amplify the targeted genomic region. We selected five genes of particular biologic interest. Three of these genes, JAG1, PDGFC, and CDKN1C were down regulated in the premalignant signature while THOC3 and LOC388796 were up regulated. For each gene, we designed primers to interrogate the methylation status of the promoter and associated CTCF binding sites. A total of 16 amplicons were evaluated with initial optimization results shown in Table 1. The primer analysis using the Agilent Bioanalyzer is shown for a subset of amplicons in Figure 8.

Amplicon:		Optimization 1		Optimization 2		Optimization 3		Optimization 4			
	Expected size	Controls	LCMs	Controls	LCMs	Controls	LCMs	Controls	LCMs	Temps.	Notes
CDKN1C-promorer 1	259	Y	N	Y	N	N	N	Y	N (PD)	58, 56, 54	Re-design
CDKN1C-promoter 2	402	Y	N	Y	N (PD)	N	N	Y	N (PD)	58, 56, 54	Re-design
CDKN1C-promoter 3	282	N	N	N	N	Y	N	Y (PD)	N (PD)	58, 60, 58-1ul, 56-1ul	Re-design
CDKN1C-CTCF	408	Y	N	Y	N	N	N	Y	N	58, 56, 54	Re-design
JAG1-promoter 1	408	Y	N (PD)	-------	-------	-------	-------	-------	-------	58	Re-design
JAG1-promoter 2	357	N	N	N	N	N	N	-------	-------	58, 56, 54	Re-design
JAG1-CTCF 1	225	Y	Y	-------	-------	-------	-------	-------	-------	58	Run at 58ºC
JAG1-CTCF 2	144	Y	Y	-------	-------	-------	-------	-------	-------	58	Run at 58ºC
PDGFC-CTCF	212	Y	Y	-------	-------	-------	-------	-------	-------	58	Run at 58ºC
PDGFC-promoter	292	N	N	N	N	Y	N (PD)	-------	-------	58, 60, 58-1ul	Re-design
LOC-CTCF 1	312	N	N	N	N	Y	N, Y	Y	N	58, 60, 58-1ul, 56-1ul	Re-design
LOC-CTCF 2	294	N	N	N	N	N	N	Y	N	58, 60, 58-1ul, 56	Re-design
LOC-5' UTR 1	391	Y	N (PD)	Y	N (PD)	N	N	-------	-------	58, 56, 54	Re-design
LOC-5' UTR 2	275	Y	Y	N	N	-------	-------	-------	-------	58, 60	Run at 58ºC
THOC3-CTCF	220	Y	Y	-------	-------	-------	-------	-------	-------	58	Run at 58ºC
THOC3-promoter	489	Y	N	-------	-------	-------	-------	-------	-------	58	Re-design

Table 1. Results of optimization using the EpiTYPER platform. Amplicons spanning the promoters and associated CTCF binding sites for five genes were evaluated. In some instances, 2 or 3 amplicons were designed to ensure interrogation of the methylation status of the entire promoter and/or CTCF binding domain. Highlighted in blue are the five amplicons that were successful in the first round of optimization and were also successful with LCM material. Generally, amplicons worked well with the control samples (no LCM captured DNAs), but failed with the laser capture samples. Successful amplicons ranged in size from 144-275bp, while unsuccessful amplicons ranged in size from 259-489bp. PD, primer dimer/non specific amplification.

Successful amplicons were then used to quantitatively determine the methylation status of individual CpGs in a small series of ovarian cancer (3) and normal samples (2). These results are shown in Figure 10.

Given the amount of effort and time required to design and optimize primer conditions, our next round of optimization and validation focused on CDNK1C. CDKN1C (p57/Kip2) is an imprinted (maternally expressed) cell cycle regulatory gene on chromosome 11p15.4. Importantly, disruption of CDKN1C expression causes the cancer predisposing syndrome Beckwith-Wiedemann. CDKN1C has also been implicated as a tumor suppressor gene in a number of human malignant neoplasms including breast, lung, pancreatic, bladder, esophageal and a variety of hematological and myeloid neoplasms. While CDKN1C dysregulation has not been extensively studied in ovarian carcinoma, the majority (75%) of sporadic ovarian carcinomas demonstrate reduced CDKN1C protein expression (<10% of tumor cells) using IHC. Thus it is important to understand the mechanism responsible for reduced expression of CDKN1C in ovarian carcinomas.

Previously data added to the UCSC Genome Browser revealed additional putative CTCF binding sites at the CDKN1C locus ((CDKN1C_01 to _04, Figure 9). Given the biological significance of the potential regulatory sites, primers were designed to allow for the evaluation of methylation status of these four CTCF binding domains as well as LOC388796 5'UTR 1 since the other amplicon LOC388796 5'UTR 2 indicated differential methylation (see Figure 10). A summary of primer optimization results in shown in Table 2.

		Optimization 1		Optimization 2		Optimization 3			
Amplicon:	Expected size	Controls	LCMs	Controls	LCMs	Controls	LCMs	Temps.	Notes
CDKN1C_01	224	N (PD)	N (PD)	N (PD)	N (PD)	-------	-------	58°, 56°	Re-design
CDKN1C_02	300	Y/N	Y	Y	Y	-------	-------	58° Good.	Run at 58º
CDKN1C_03	335	Y	Y	-------	-------	-------	-------	58° Good.	Run at 58º
CDKN1C_04	222	Y/N (PD)	N (PD)	Y (PD)	Y/N (PD)	Y/N (PD)	Y/N (PD)	58°, 60°, 62°	Re-design
LOC-5' UTR 1	190	Y	Y/N	Y	Y	-------	-------	58°, 56° Good.	Run at 56º

Table 2. Results of optimization using the EpiTYPER platform. Amplicons spanning the four putative CDKN1C-associated CTCF binding sites as well as the 5'UTR of LOC388796 were evaluated. Highlighted in blue are the three amplicons that were successful in this round of optimization which were also successful with LCM material. PD, primer dimer/non specific amplification.

Our survey of cytosine methylation in a small subset of normal and tumor tissues indicates an increase in DNA methylation at the CDKN1C tumor suppressor gene and the 5'UTR of LOC388796 (see Figure 11). To confirm the significance of this observation, we prepared DNA from an additional 20 microdissected ovarian tumors and 5 normal fallopian tubes. The Krumm lab requested and was granted a no-cost extension to allow for completion of the analysis on these samples. Results are shown in Figure 12. While initial results indicated that methylation of CDKN1C may contribute to down regulation in tumors, our results indicate that it is not the primary mechanism.

During this period we became aware of a manuscript describing loss of CDKN1C in sporadic breast cancers. Rodriquez et al, reported that in breast cancer cell lines epigenetic silencing of CDKN1C occurs in part as the result of genetic loss of the inactive methylated allele (6). They also identified a novel cis-encoded antisense transcript, CDKN1C-AS, which is induced by estrogen following pharmacologic inhibition of DNA methyltransferase and histone deacetylase activity. When overexpressed, CDKN1C-AS was capable of repressing endogenous CDKN1C *in vivo* suggesting that in addition to promoter hypermethylation, epigenetic repression of tumor suppressor genes by CTCF and noncoding RNA transcripts could be more common and important than previously understood.

In order to determine whether other mechanisms contribute to reduced CDKN1C expression observed in our ovarian studies we determined whether in fallopian tube epithelium CDKN1C-AS antisense message regulated expression of CDKN1C in response to estrogen (data not shown). Our results indicate that unlike those of Rodriquez et al., in breast epithelium, CDKN1C-AS does not regulate CDKN1C expression in FT epithelium. In addition we determined whether LOH of the CDKN1C contributed to loss of CDKN1C in tumors (data not shown). Again LOH does not contribute to loss of CDKN1C in tumors. While our results contradict those of Rodriquez et al., we anticipate that our data will be submitted as a manuscript in early 2012. Importantly our results indicate that as yet to be discovered mechanism(s) regulate expression of CDKN1C in normal FT epithelium.

KEY ACCOMPLISHMENTS

- We identified and confirmed 41 probe sets which constitute a *BRCA1* preneoplastic profile which includes several known tumor suppressors and oncogenes.

- Expression of a subset of genes was validated with quantitative reverse transcription-polymerase chain reaction and immunohistochemistry.

- Our epigenetic studies further refined and confirmed evidence for antagonistic action of DNA methylation and CTCF binding at several loci.

- We have demonstrated that for some binding sites, CTCF requires methylation of very specific cytosine residues within the target motif.

- To quantitatively define changes in cytosine methylation in genes that constitute a premalignant signature for ovarian cancer, we established the experimental protocol for the EpiTYPER analyses on LCM DNA.

- EpiTYPER analysis indicated that methylation of CDKN1C occurs in a small fraction of ovarian cancers.

- Reduced CDKN1C expression in histologically normal FT epithelium in BRCA1 mutation carriers does not appear to be the result of methylation of CTCF binding sites, estrogen induced expression of CDKN1C-AS nor LOH.

REPORTABLE OUTCOMES

Press JZ, Wurz K, Norquist BM, Lee MK, Pennil C, Garcia R, Welcsh P, Goff BA, **Swisher EM**. *Identification of a preneoplastic gene expression profile in tubal epithelium of BRCA1 mutation carriers.* **Neoplasia.** 2010 Dec;12(12):993-1002.

Thomas BJ, Rubio ED, Krumm N, Broin PO, Bomsztyk K, Welcsh P, Greally JM, Golden AA, **Krumm A**. *Allele-specific transcriptional elongation regulates monoallelic expression of the IGF2BP1 gene.* **Epigenetics & Chromatin.** 2011 Aug 3;4:14.

CONCLUSION

By comparing the histologically normal FTE from women who carry BRCA1 mutations and have micro-invasive foci with normal FTE from women without BRCA1 mutations, we have identified a premalignant expression signature which may reflect early steps in BRCA1-mediated ovarian carcinogenesis. The significant overlap in genes differentially expressed between BRCA1 normal FTE and BRCA1 ovarian carcinomas confirms that our relatively high risk approach has paid off. We have used histologically normal BRCA1 FT near an identifiable neoplastic FT lesion to identify alterations in gene expression profiles that contain the same as expression differences in BRCA1 ovarian carcinomas.

Our preliminary epigenetic analyses did not detect any differences between BRCA1 cancers and normal risk FTE for JAG1 at both CTCF sites 1 and 2, the THOC3 CTCF site, the PDGFC CTCF site and amplicon 1 of the 5'UTR of LOC388796 (amplicon 1; Figure 10). However, our initial quantitative methylation analyses indicate differences in methylation at amplicon 2 of 5'UTR of LOC388796 (amplicon 2; Figure 10) and two of the CTCF sites associated with CDNK1C (CDNK1C-02, and -03; Figure 11).

Given the potential biological impact of loss of CDKN1C expression in ovarian carcinomas and our finding that in some the mechanism responsible for reduced expression may be methylation of associated CTCF binding domains, we optimized primers spanning CTCF site CDKN1C-01 (Figure 9). We requested and were granted additional time to allow for required primer design and PCR optimization for CDKN1C-associated CTCF binding domains. We performed detailed methylation mapping of CDKN1C CTCF sites in an additional 40 microdissected samples including 15 from fallopian tubal (FT) epithelium (7 normal, 7 FT occult cancer, 1 BRCA1+ ovarian cancer), and 25 from ovarian epithelium (8 BRCA1+ ovarian cancer, 5 BRCA2+ ovarian cancer and 12 sporadic ovarian cancer). Because our methylation analysis did not indicate that loss of CDKN1C expression was due to methylation of CTCF binding domains in the CDKN1C locus, we evaluated whether other mechanisms contributed to loss of CDKN1C in the ovarian premalignant gene expression profile.

REFERENCES:

1. Witcher, M. and B.M. Emerson, *Epigenetic silencing of the p16(INK4a) tumor suppressor is associated with loss of CTCF binding and a chromatin boundary.* Mol Cell, 2009. **34**(3): p. 271-84.
2. Phillips, J.E. and V.G. Corces, *CTCF: master weaver of the genome.* Cell, 2009. **137**(7): p. 1194-211.
3. Gombert, W.M. and A. Krumm, *Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression.* PLoS One, 2009. **4**(7): p. e6109.
4. Hark, A.T., et al., *CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus.* Nature, 2000. **405**(6785): p. 486-9.
5. Rubio, E.D., et al., *CTCF physically links cohesin to chromatin.* Proc Natl Acad Sci U S A, 2008. **105**(24): p. 8309-14.

6. Rodriguez BA, Weng YI, Liu TM, Zuo T, Hsu PY, Lin CH, Cheng AL, Cui H, Yan PS, Huang TH. *Estrogen-mediated epigenetic repression of the imprinted gene cyclin-dependent kinase inhibitor 1C in breast cancer cells.* Carcinogenesis. 2011. 32(6):812-21.

SUPPORTING DATA

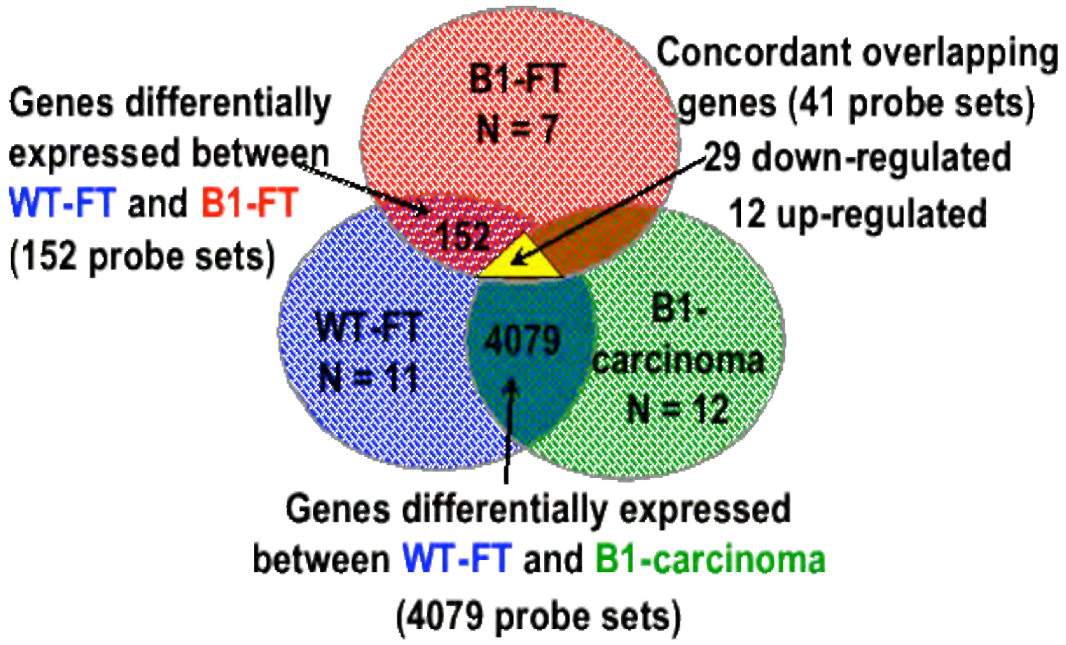


Figure 1. Diagram illustrating the protocol used to define the gene signature: Pairwise comparison between the WT-FT group and the B1-FT group (fold change ≥1.8, p-value <0.01) identified 152 differentially expressed probe sets. Pairwise comparison between the WT-FT group and the B1-carcinoma group (fold change ≥1.8, p-value <0.01) identified 4079 differentially expressed probe sets. To minimize the false discovery rate probe sets were only included in the gene signature with concordant down-regulation or up-regulation in both pairwise comparisons. The 41 probe sets fulfilling this criteria are shown in the Table 1.

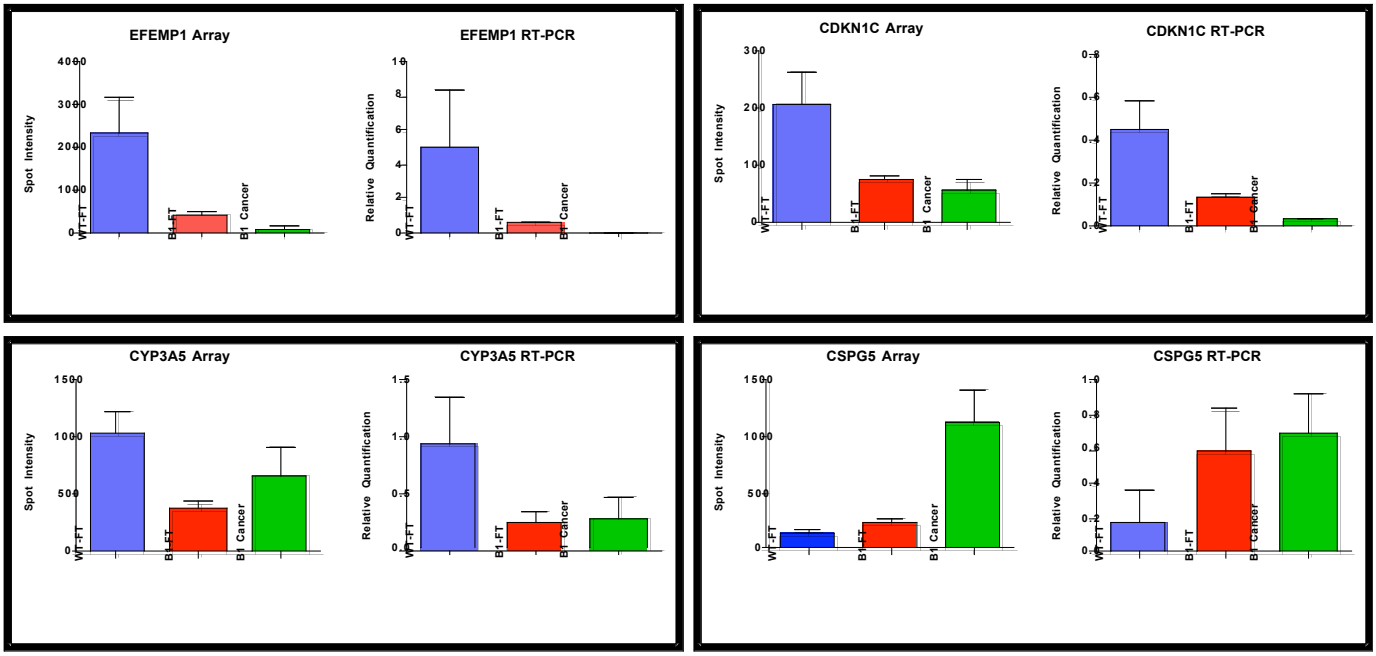


Figure 2. Real-time RT-PCR results: Four genes from the gene signature were selected for validation by RT-PCR with Taqman assays, using five cases from each group. EFEMP1 (EGF-containing fibulin-like extracellular matrix protein 1), CDKN1C (Cyclin-dependent kinase inhibitor 1C or p57), CYP3A5 (Cytochrome p450, family 3, subfamily A), and CSPG5 (Chondroitin sulfate proteoglycan 5 or neuroglycan C). Array corresponded well with RT-PCR results as shown.

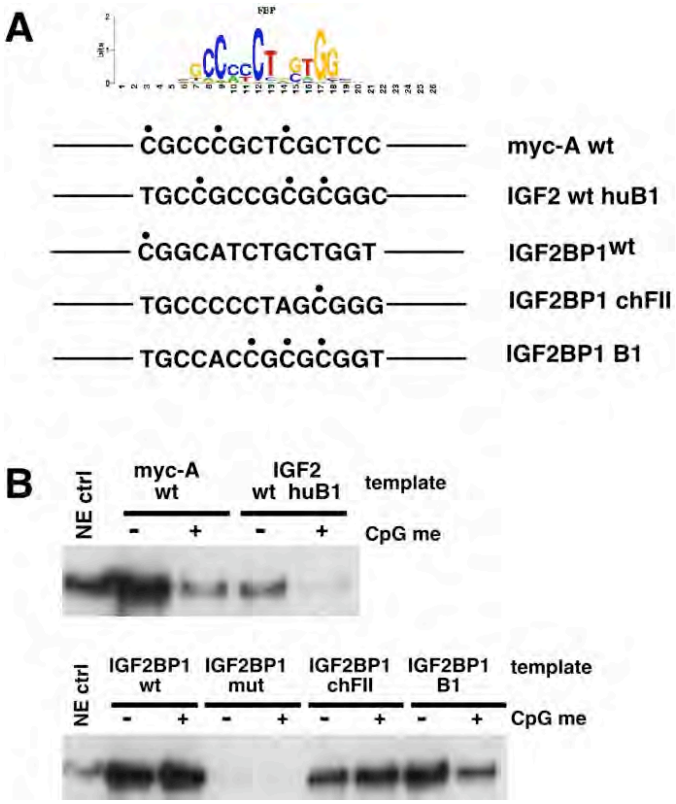


Figure 3: Methylation sensitive binding of CTCF to its target sequences. (A) CTCF target motifs derived from the MYC, IGF2, and IGF2BP1 gene loci were tested for their ability to bind CTCF in their unmethylated and methylated form. Methylable cytosine residues are indicated by filled circles. (B) Western blots reveal the amount of CTCF bound to the target motifs. Binding of CTCF to myc-A and IGF2 huB1 is significantly inhibited by methylation (+ CpG me). In contrast, methylation of the IGF2BP1 target motif does not affect recruitment of CTCF (compare IGF2BP1wt +/- CpGme). Modification of the IGF2BP1 target sequence into a motif that resembles that of the IGF2 gene (IGF2BP1 B1) restores the methylation sensitivity of CTCF binding. NE ctrl represents signal obtained from nuclear extract.

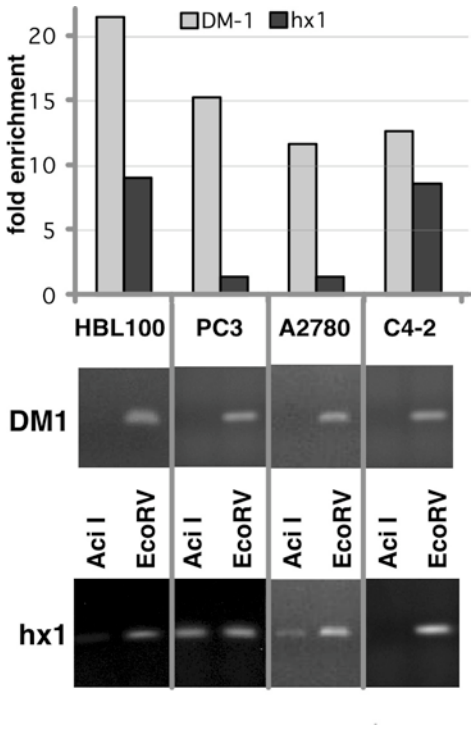


Figure 4. DNA methylation-sensitive binding of CTCF at the HoxA locus. Upper panel, enrichment of CTCF bound sequences at the DM1 locus (grey bars) and the hx1 binding site (black bars) in ChIP in breast epithelial cell type HBL100, ovarian cancer cell line A2780, and prostate cancer cell lines PC3 and C4-2. While CTCF is bound to the DM1 locus in all cell lines (grey bars), it fails to bind to the hx1 site at the HoxA gene domain in PC3 and A2780 (black bars). Lower panel, DNA methylation analysis by PCR of genomic DNA after restriction digest with methylation-sensitive Aci I or Eco RV (control digest; amplicons do not contain Eco RV sites) reveal no DNA methylation at DM1 (no PCR product due to digest of DNA). In contrast, hx1 in PC3 and A2780 is methylated, leading to inhibition of CTCF binding.

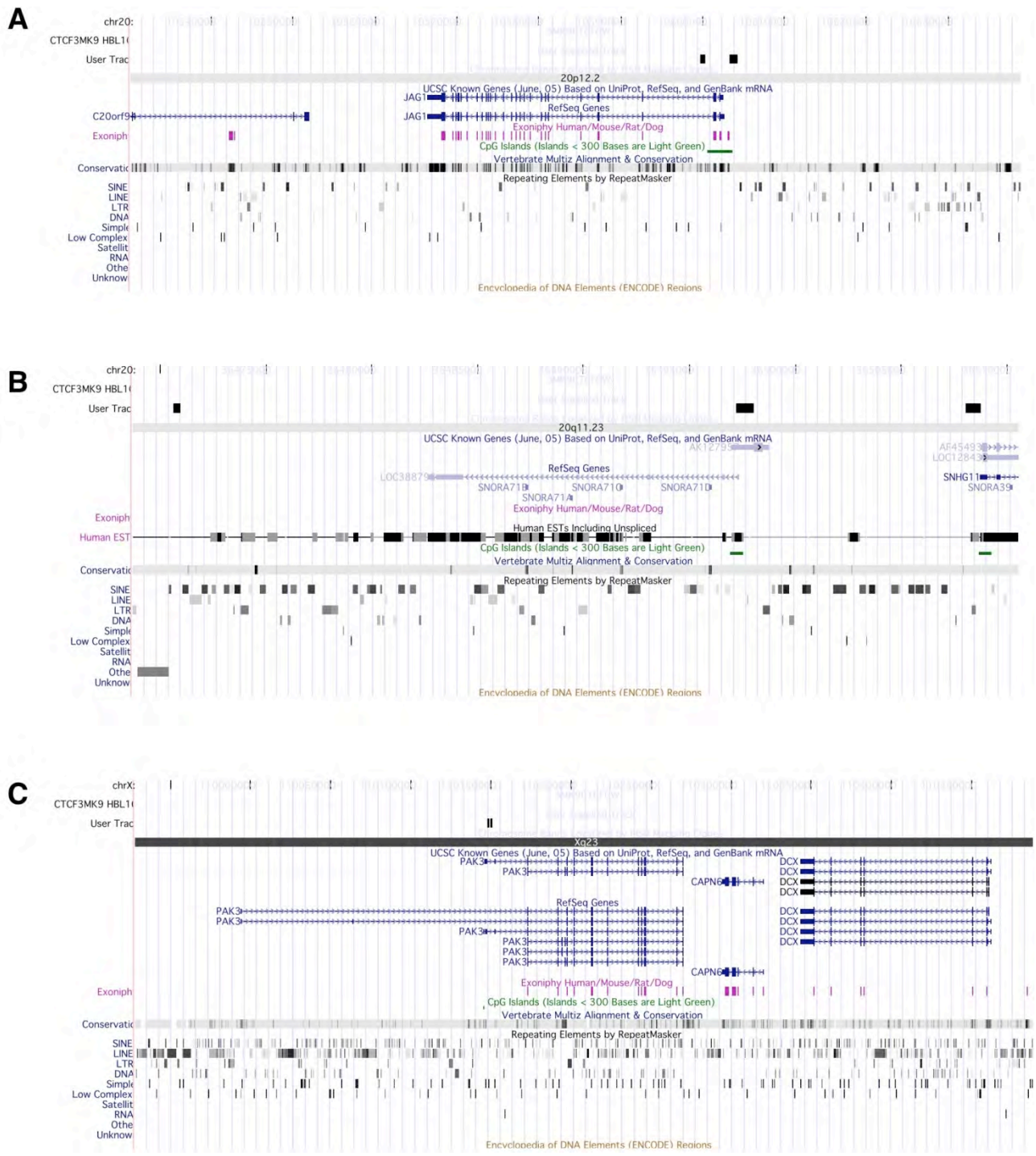


Figure 5. Distribution of CTCF sites at genomic loci that are differentially expressed and part of the premalignant gene signature. Custom tracks of the UCSC genome browser of the JAG1 (A), LOC38879 (B), and PAK3 (C) genes are shown as examples. CTCF binding regions are indicated by black boxes within "User track". Chromosome and position are shown at the top.

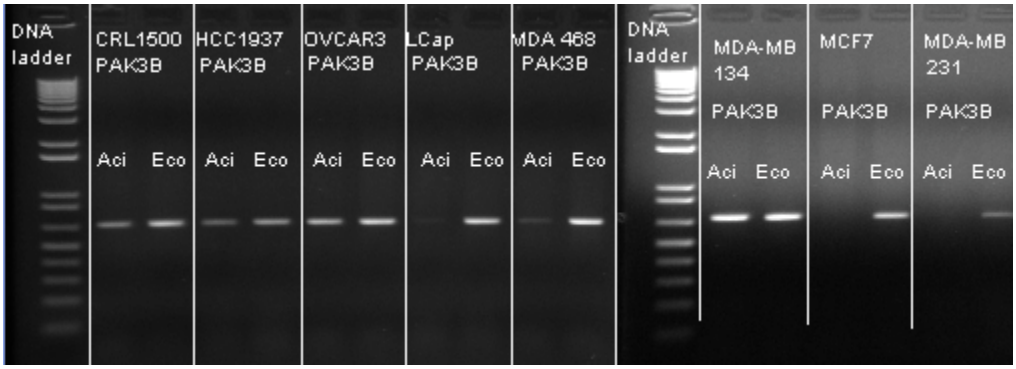


Figure 6. Differential methylation of CTCF binding sites at the PAK3 gene locus in several cancer cell lines. Genomic DNA isolated from indicated cell types was digested with AciI (CpG methylation sensitive) or EcoRV (control). Genomic region of PAK3A gene was subsequently amplified by PCR. CpG methylation at CTCF binding region blocks digest by AciI, and permits amplification of PAK3B region (e.g. ovarian cancer cell line OVCAR3 and breast cancer cell line MDA-MB134). In contrast, non-methylated PAK3 regions are digested in the presence of AciI, and PCR amplification does not yield any product (e.g. prostate cancer cell line LnCaP).

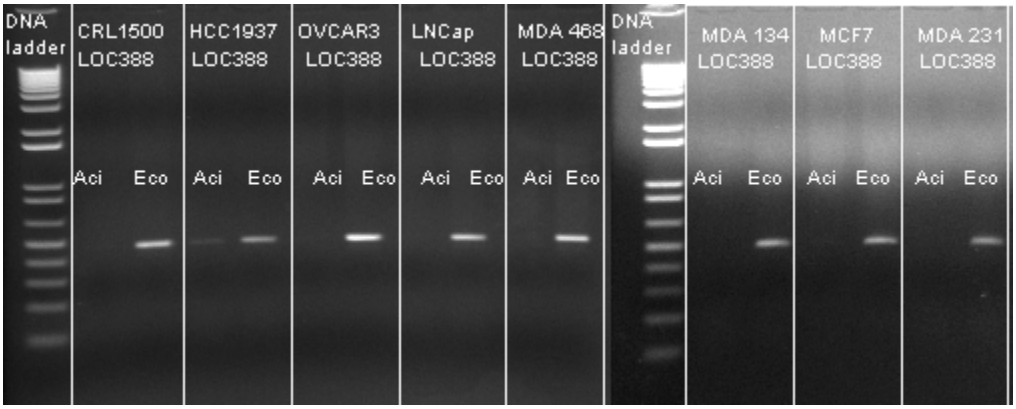


Figure 7. CTCF binding region at LOC388796 is invariably unmethylated in different tissues/cell lines (CRL1500, HCC1937, OVCAR3, LnCaP, MDA468, MDA-MB134, MCF7, and MDA-MB231). Genomic DNA digested with either AciI or EcoRV was amplified with primers specific for a CTCF binding region at LOC388796 (chr20). Absence of PCR product using genomic DNA digested by AciI indicates absence of methylation in all cell lines.

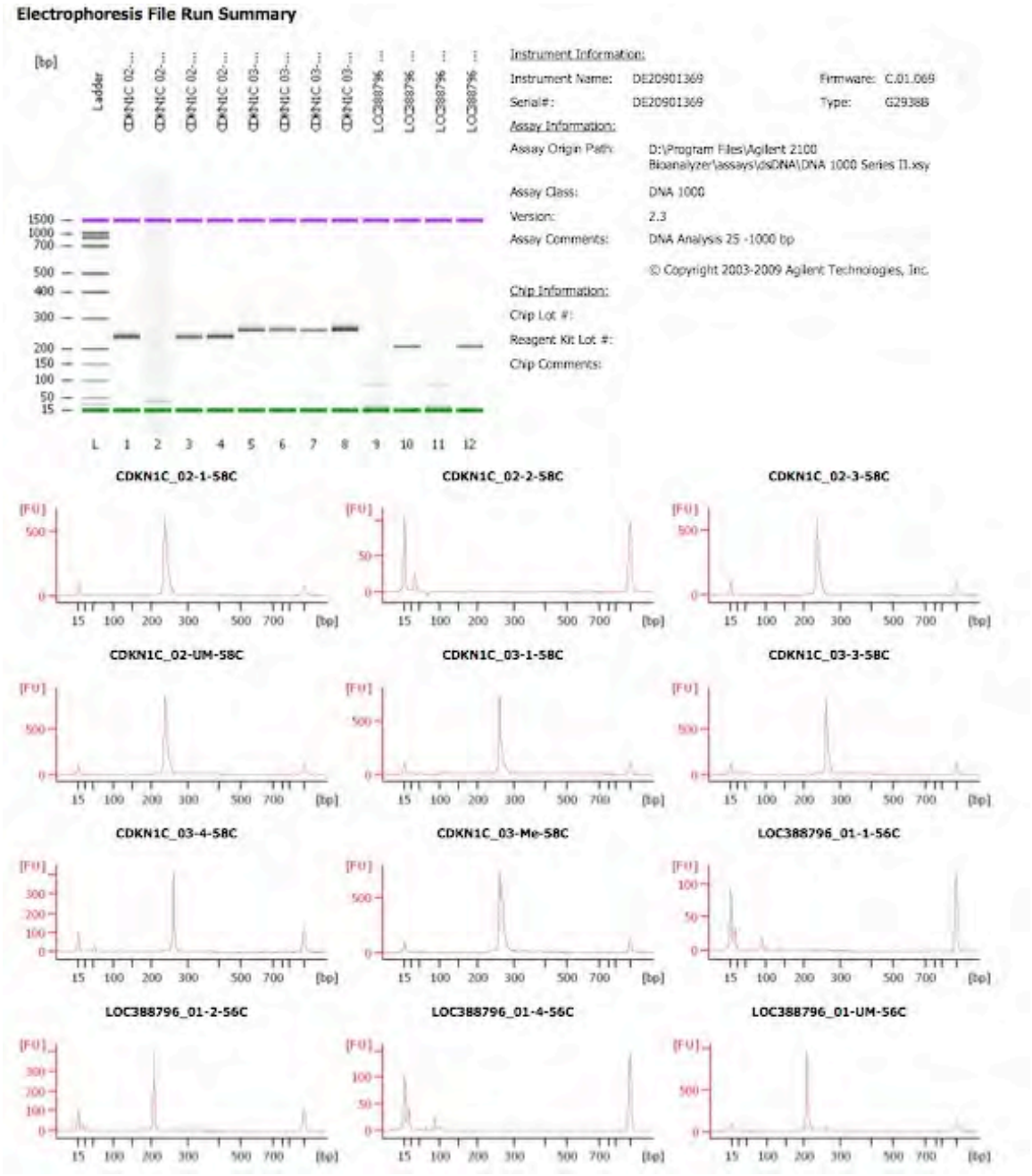


Figure 8: Example of primer optimization and analysis with the Bioanalyzer to identify primer combinations suitable for EpiTYPER assay. Several primer combinations were used to amplify genomic regions by PCR. Amplicons were analyzed using the Agilent 2100 Bioanalyzer. Top panel, electrophoresis of amplicons. Position of references are indicated by purple and green bands. Quantitative analyses of each of the PCR products is shown below.

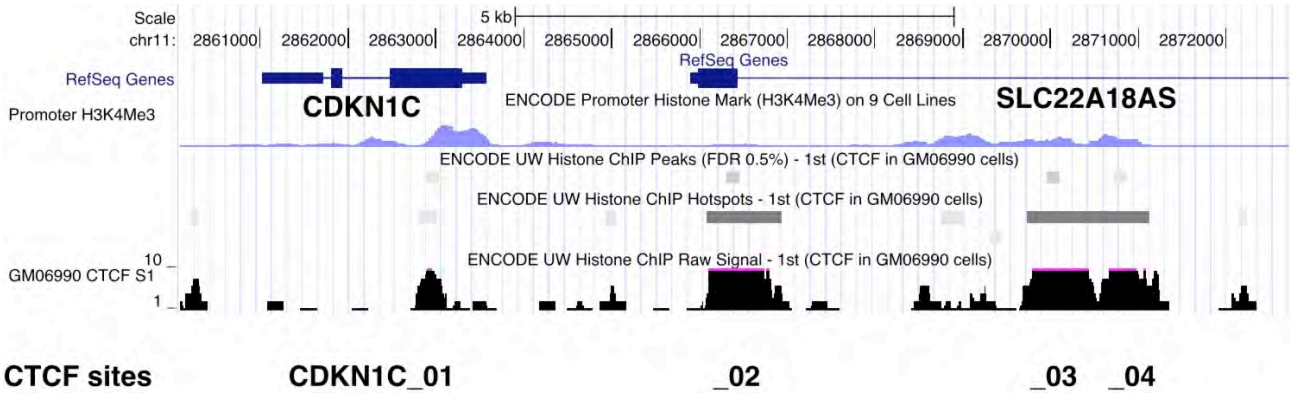
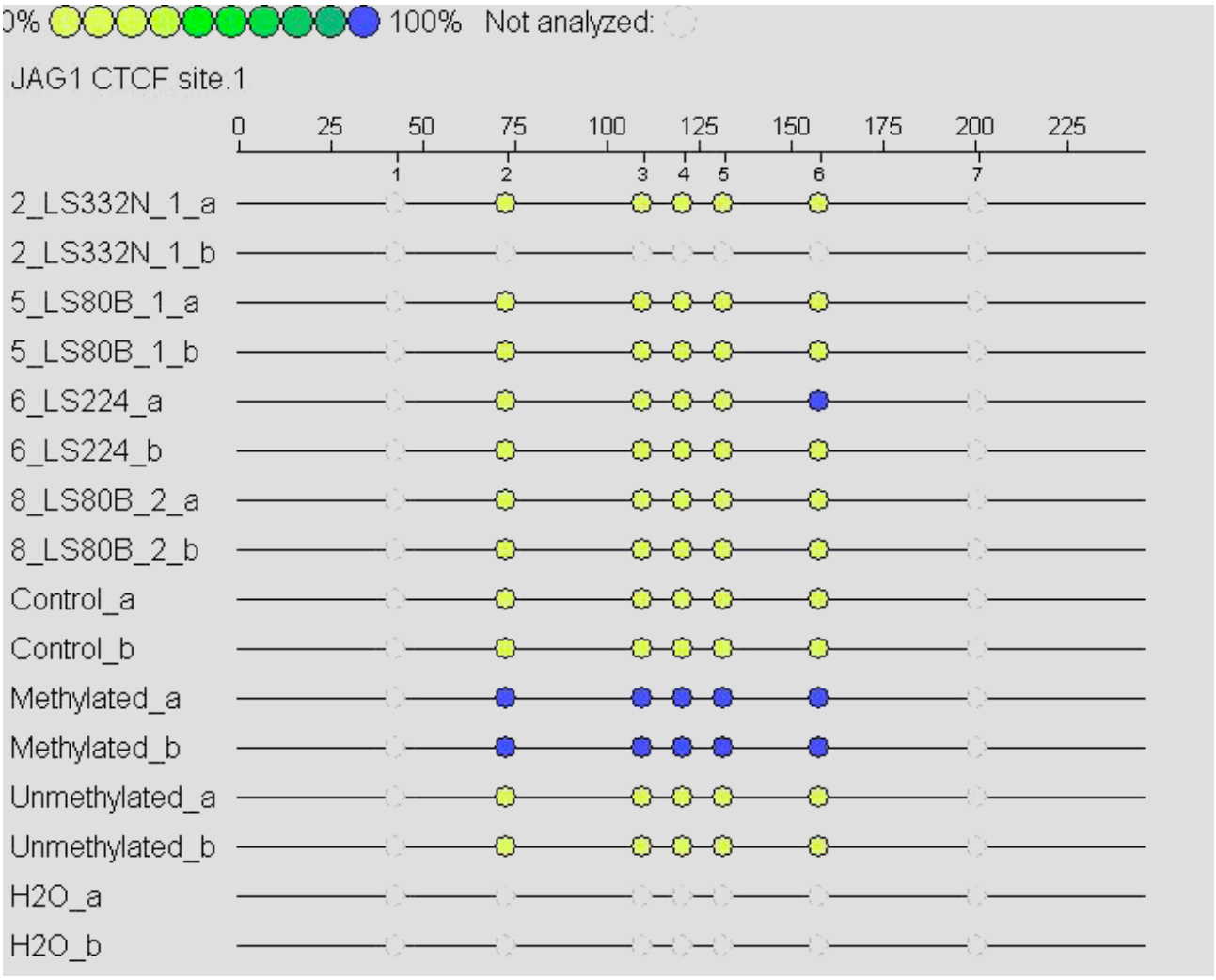
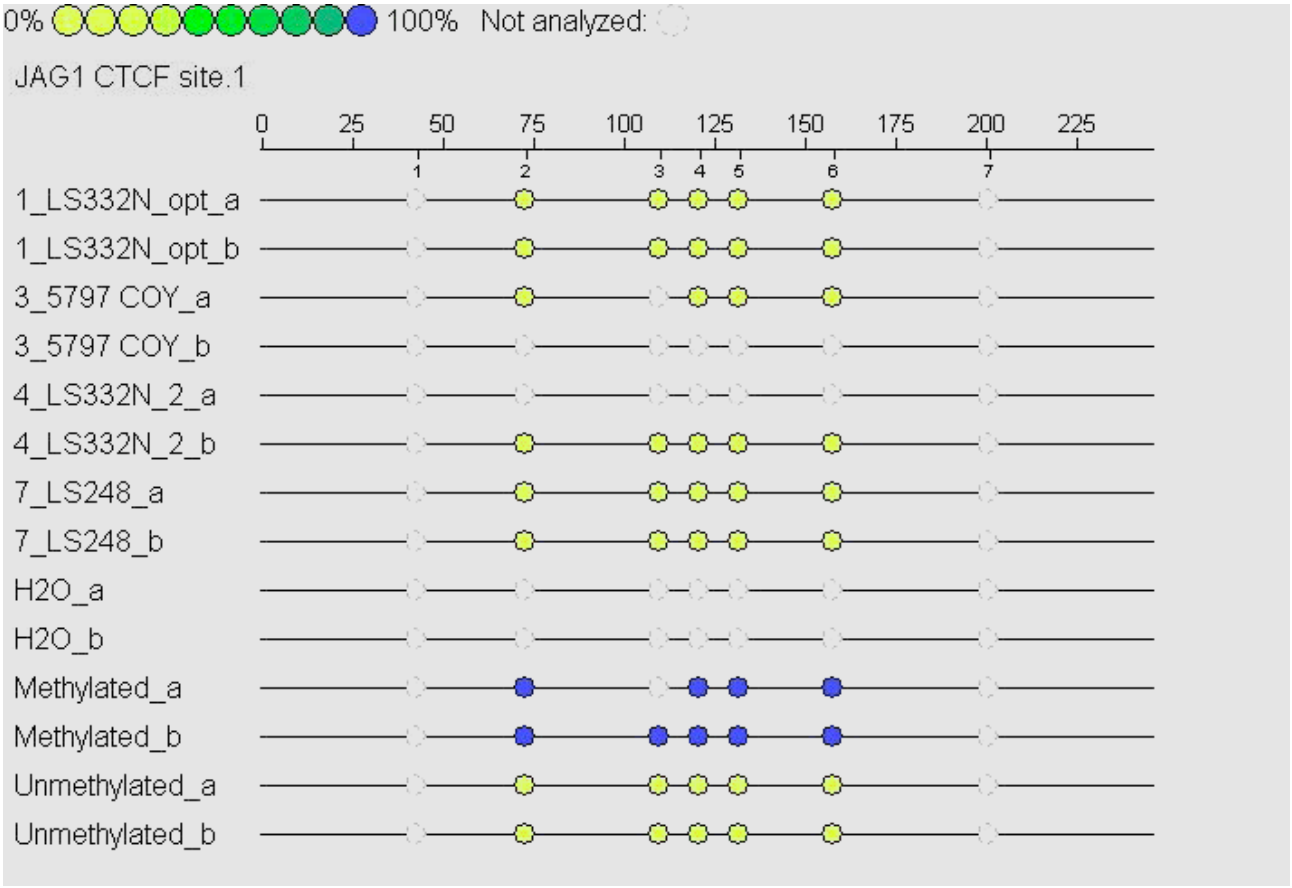
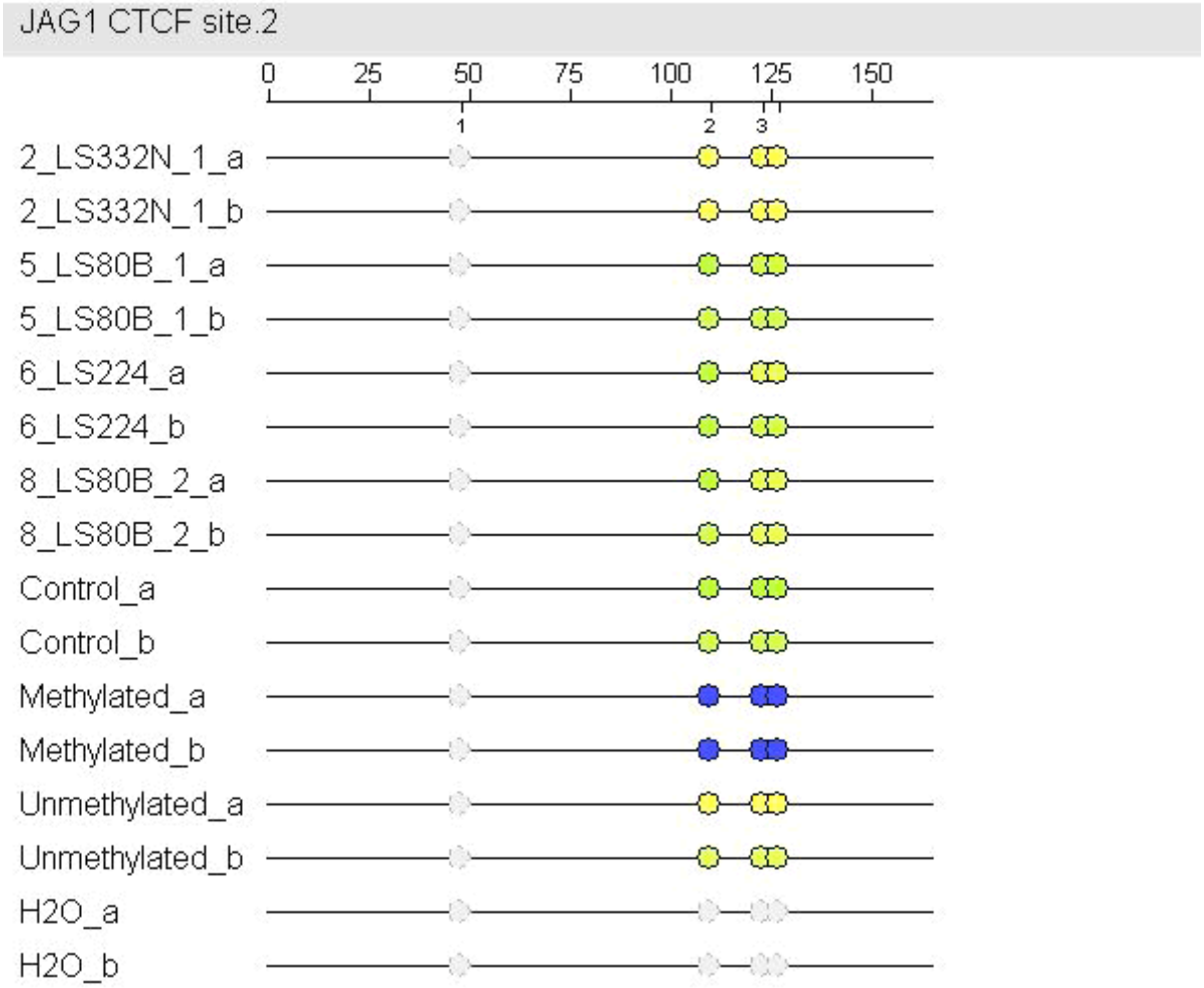
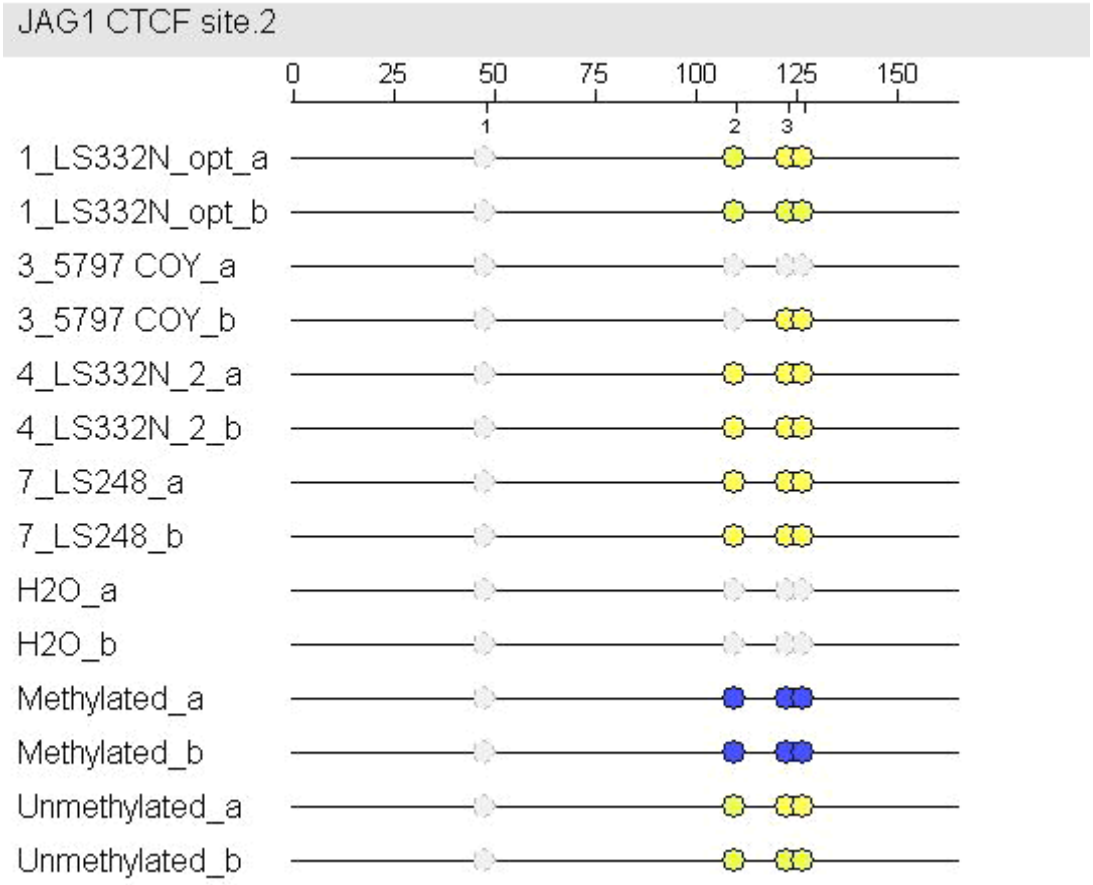


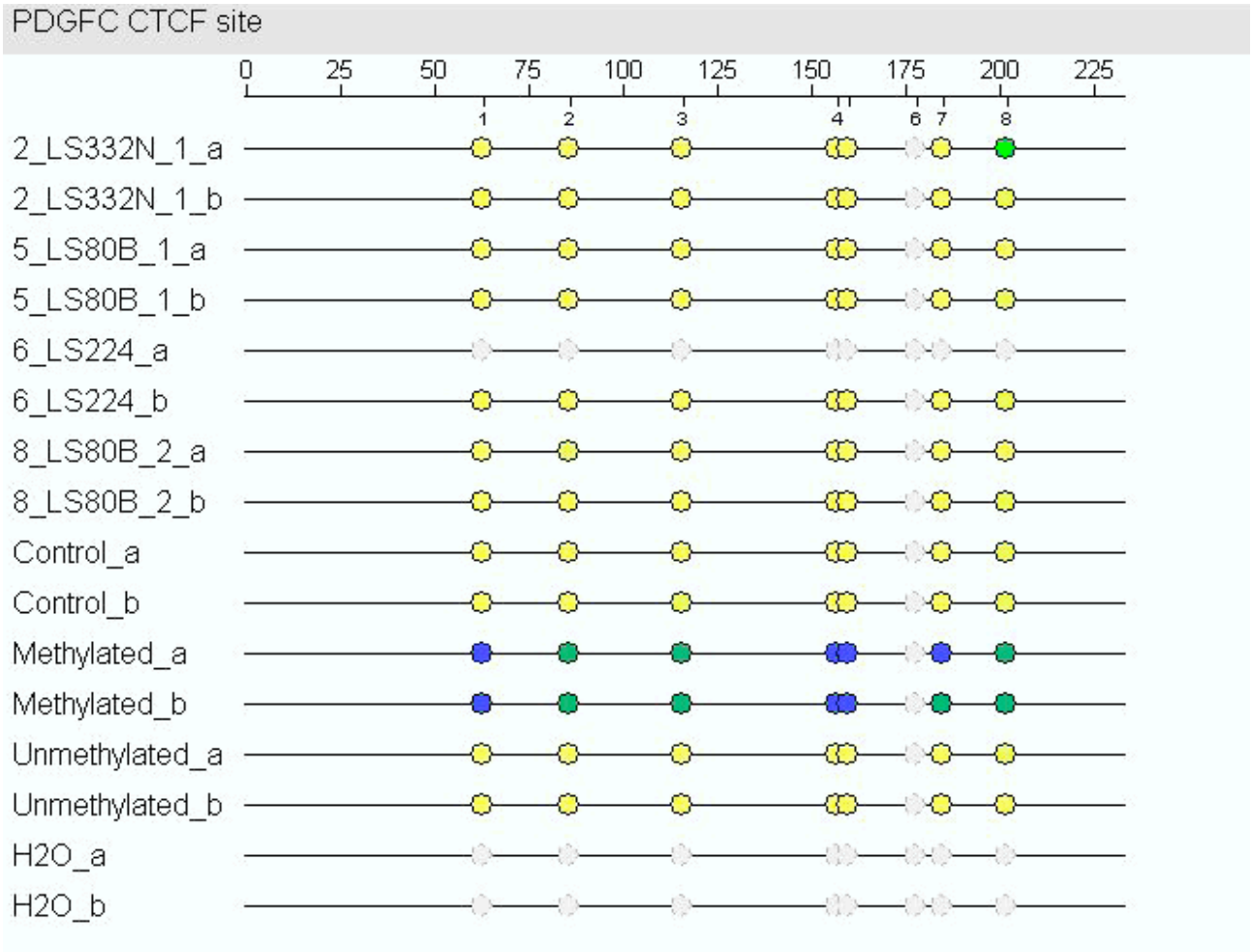
Figure 9: Distribution of CTCF sites across the CDKN1C genomic region on chromosome 11. Position of CTCF sites 1 to 3 in the GM6990 cell line relative to the position of the CDKN1C and SLC22A18AS genes are shown.

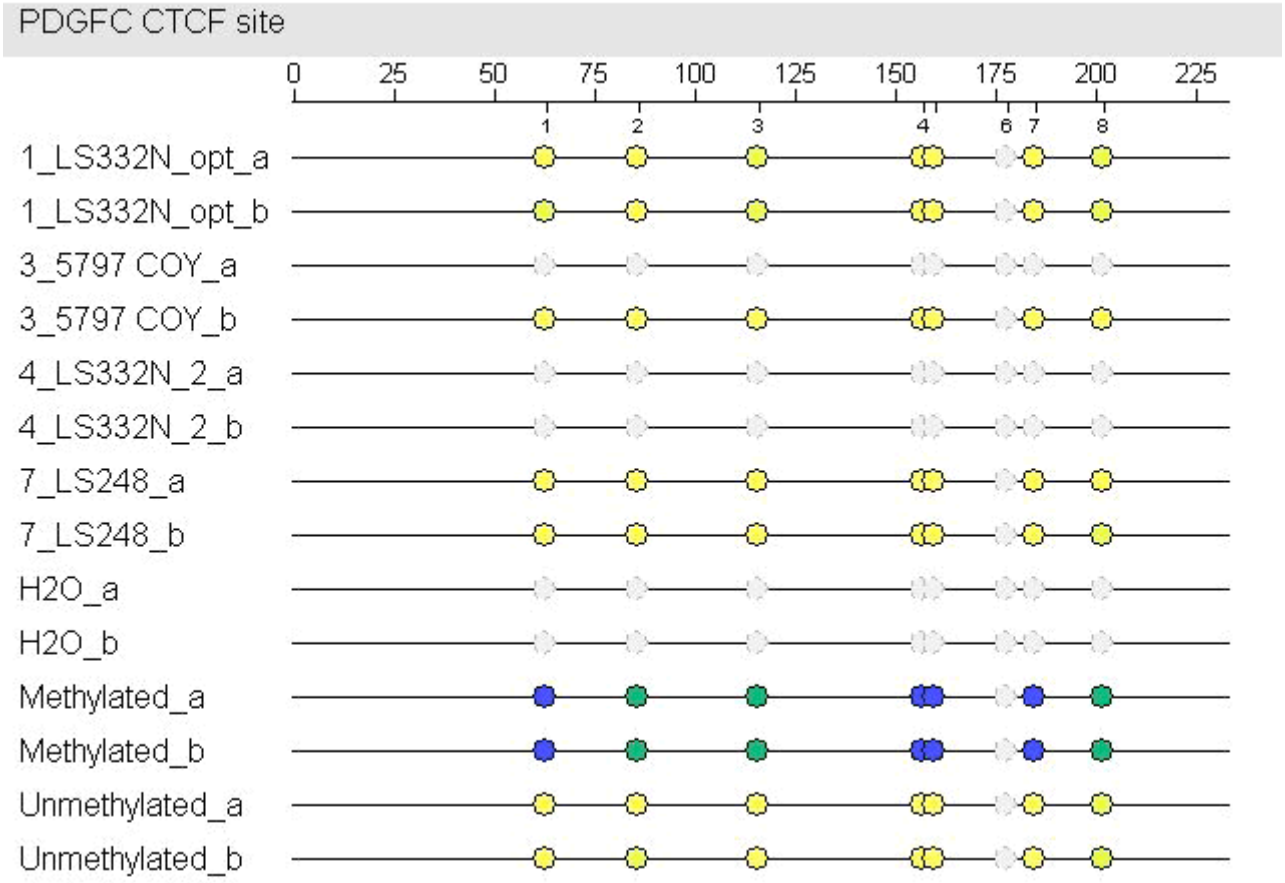

0%
100%
Not analyzed:
JAG1 CTCF site.1
0
25
50
75
100
125
150
175
200
225
1
2
3
4
5
6
7
2_LS332N_1_a
2_LS332N_1_b
5_LS80B_1_a
5_LS80B_1_b
6_LS224_a
6_LS224_b
8_LS80B_2_a
8_LS80B_2_b
Control_a
Control_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b
H2O_a
H2O_b

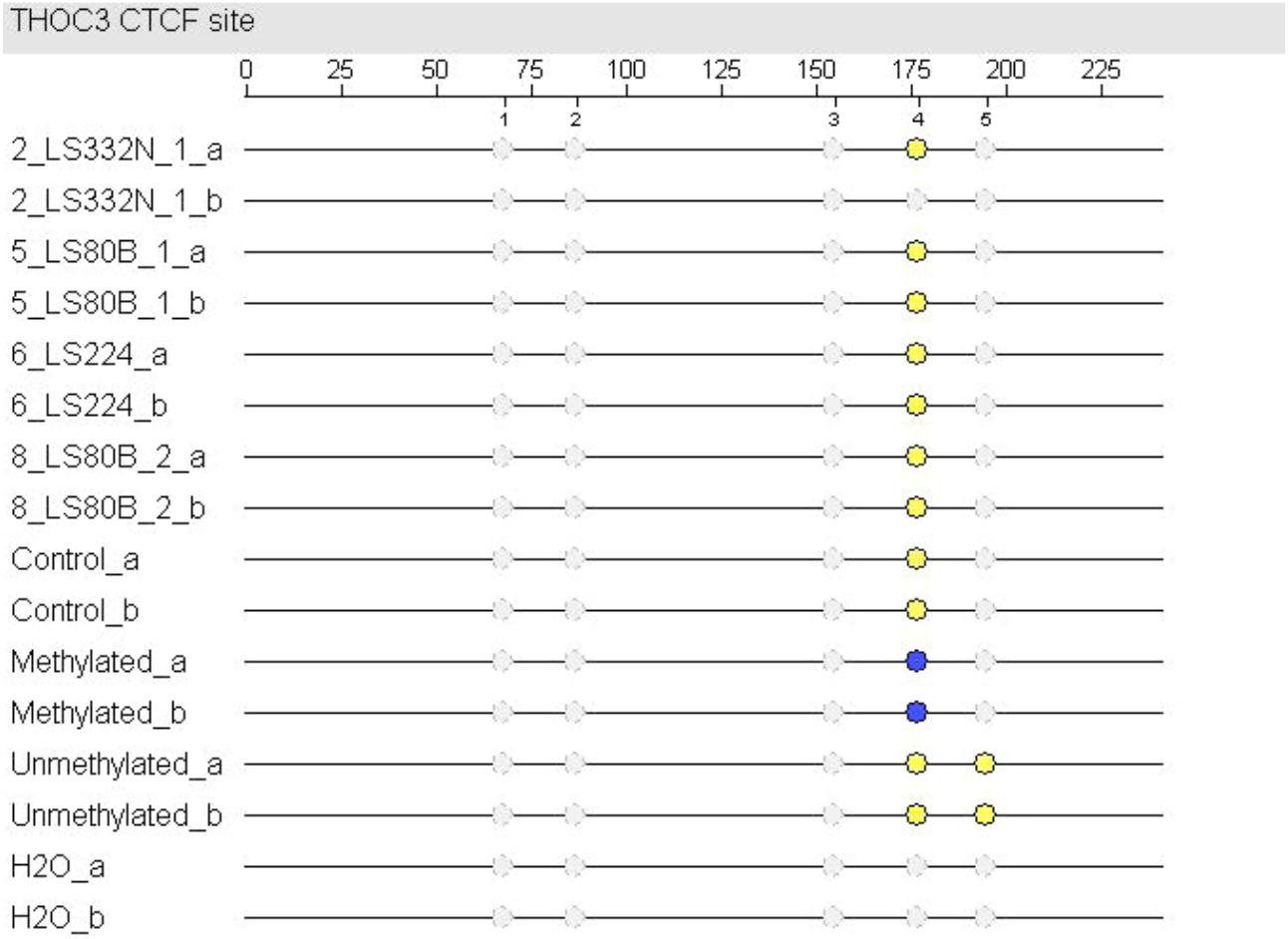

0%
100%
Not analyzed:
JAG1 CTCF site.1
0
25
50
75
100
125
150
175
200
225
1
2
3
4
5
6
7
1_LS332N_opt_a
1_LS332N_opt_b
3_5797 COY_a
3_5797 COY_b
4_LS332N_2_a
4_LS332N_2_b
7_LS248_a
7_LS248_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b

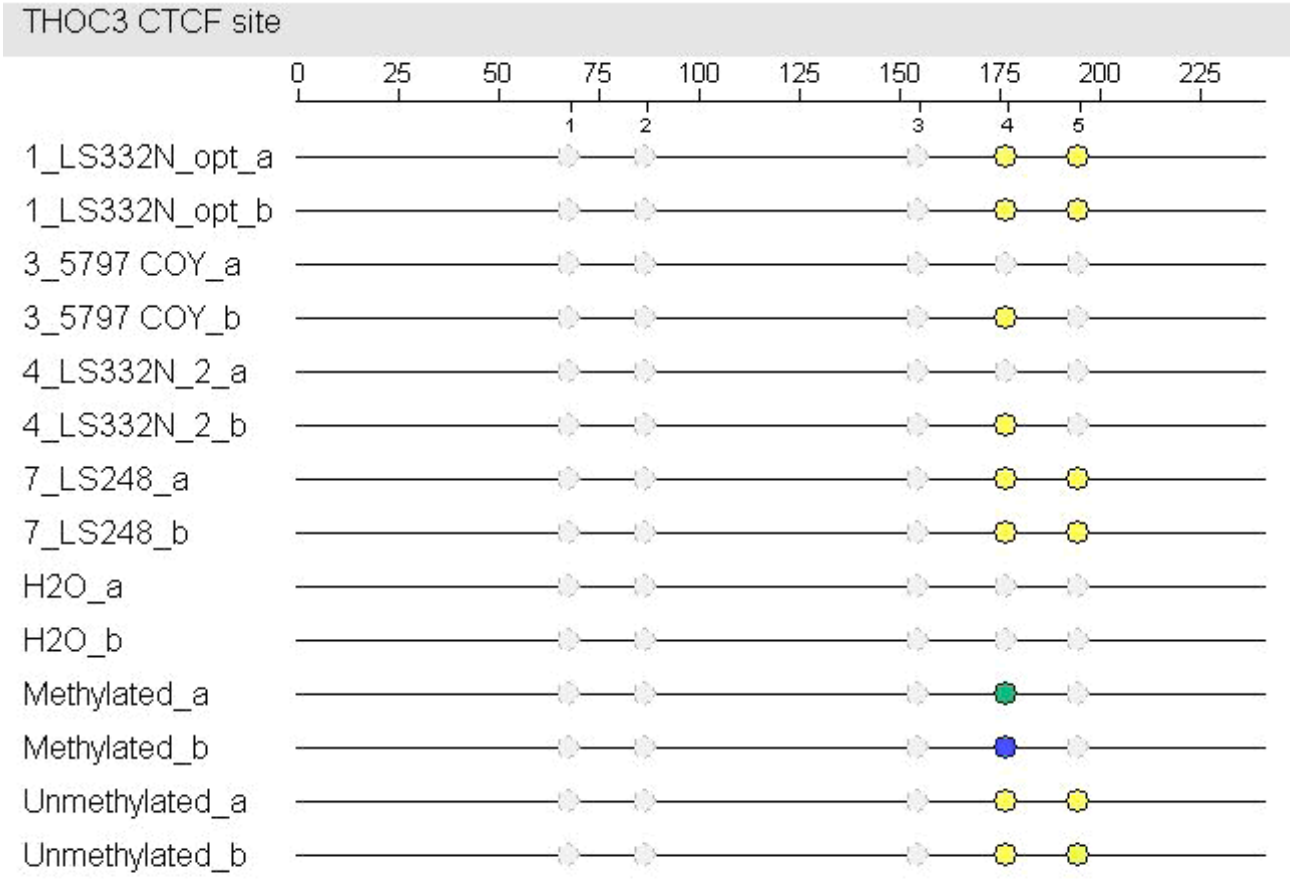

JAG1 CTCF site.2
0
25
50
75
100
125
150
1
2
3
2_LS332N_1_a
2_LS332N_1_b
5_LS80B_1_a
5_LS80B_1_b
6_LS224_a
6_LS224_b
8_LS80B_2_a
8_LS80B_2_b
Control_a
Control_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b
H2O_a
H2O_b

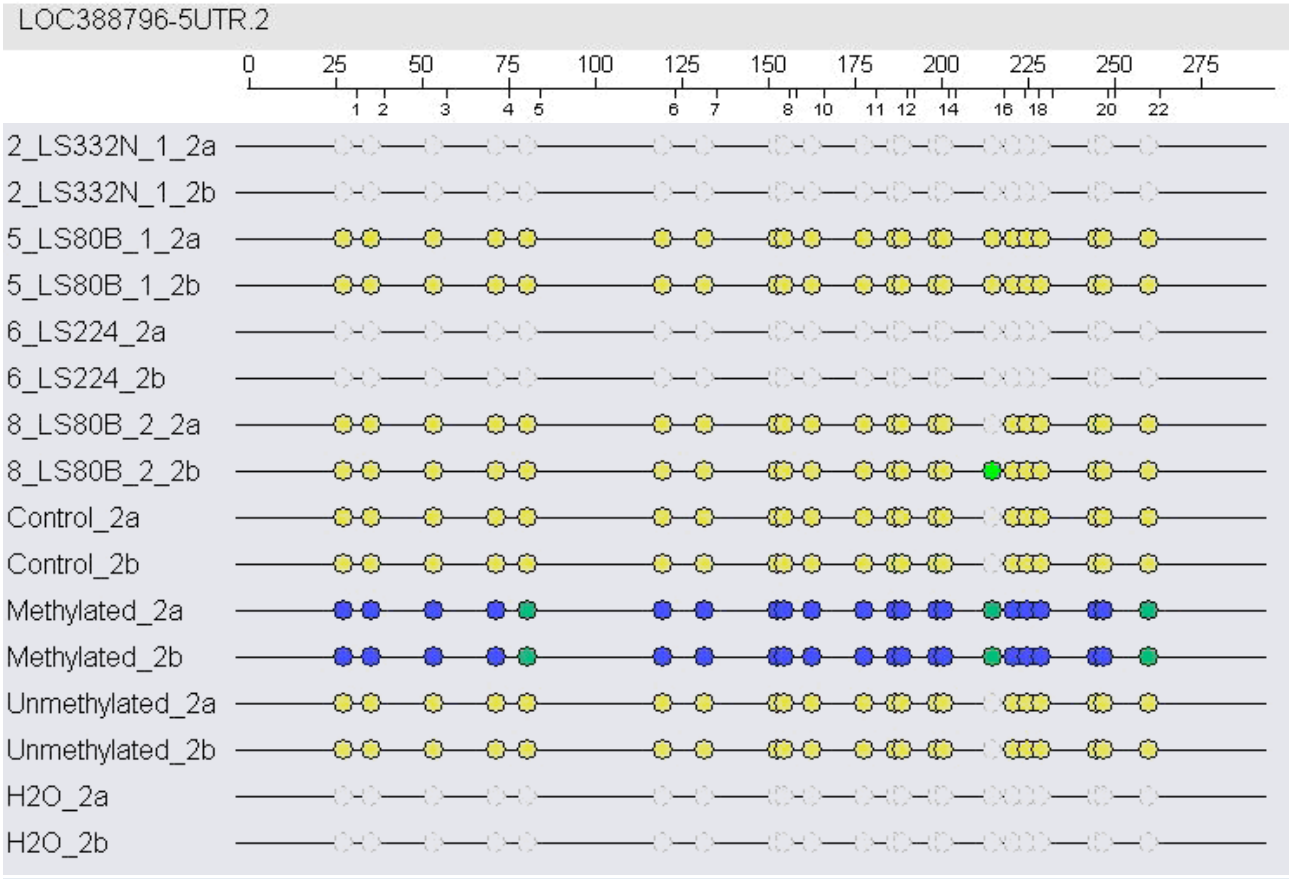

JAG1 CTCF site.2
0
25
50
75
100
125
150
1
2
3
1_LS332N_opt_a
1_LS332N_opt_b
3_5797 COY_a
3_5797 COY_b
4_LS332N_2_a
4_LS332N_2_b
7_LS248_a
7_LS248_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b


PDGFC CTCF site
0
25
50
75
100
125
150
175
200
225
1
2
3
4
6
7
8
2_LS332N_1_a
2_LS332N_1_b
5_LS80B_1_a
5_LS80B_1_b
6_LS224_a
6_LS224_b
8_LS80B_2_a
8_LS80B_2_b
Control_a
Control_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b
H2O_a
H2O_b


PDGFC CTCF site
0
25
50
75
100
125
150
175
200
225
1
2
3
4
6
7
8
1_LS332N_opt_a
1_LS332N_opt_b
3_5797 COY_a
3_5797 COY_b
4_LS332N_2_a
4_LS332N_2_b
7_LS248_a
7_LS248_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b


THOC3 CTCF site
0
25
50
75
100
125
150
175
200
225
1
2
3
4
5
2_LS332N_1_a
2_LS332N_1_b
5_LS80B_1_a
5_LS80B_1_b
6_LS224_a
6_LS224_b
8_LS80B_2_a
8_LS80B_2_b
Control_a
Control_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b
H2O_a
H2O_b


THOC3 CTCF site
0
25
50
75
100
125
150
175
200
225
1
2
3
4
5
1_LS332N_opt_a
1_LS332N_opt_b
3_5797 COY_a
3_5797 COY_b
4_LS332N_2_a
4_LS332N_2_b
7_LS248_a
7_LS248_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b


LOC388796-5UTR.2
0 25 50 75 100 125 150 175 200 225 250 275
1 2 3 4 5 6 7 8 10 11 12 14 16 18 20 22
2_LS332N_1_2a
2_LS332N_1_2b
5_LS80B_1_2a
5_LS80B_1_2b
6_LS224_2a
6_LS224_2b
8_LS80B_2_2a
8_LS80B_2_2b
Control_2a
Control_2b
Methylated_2a
Methylated_2b
Unmethylated_2a
Unmethylated_2b
H2O_2a
H2O_2b

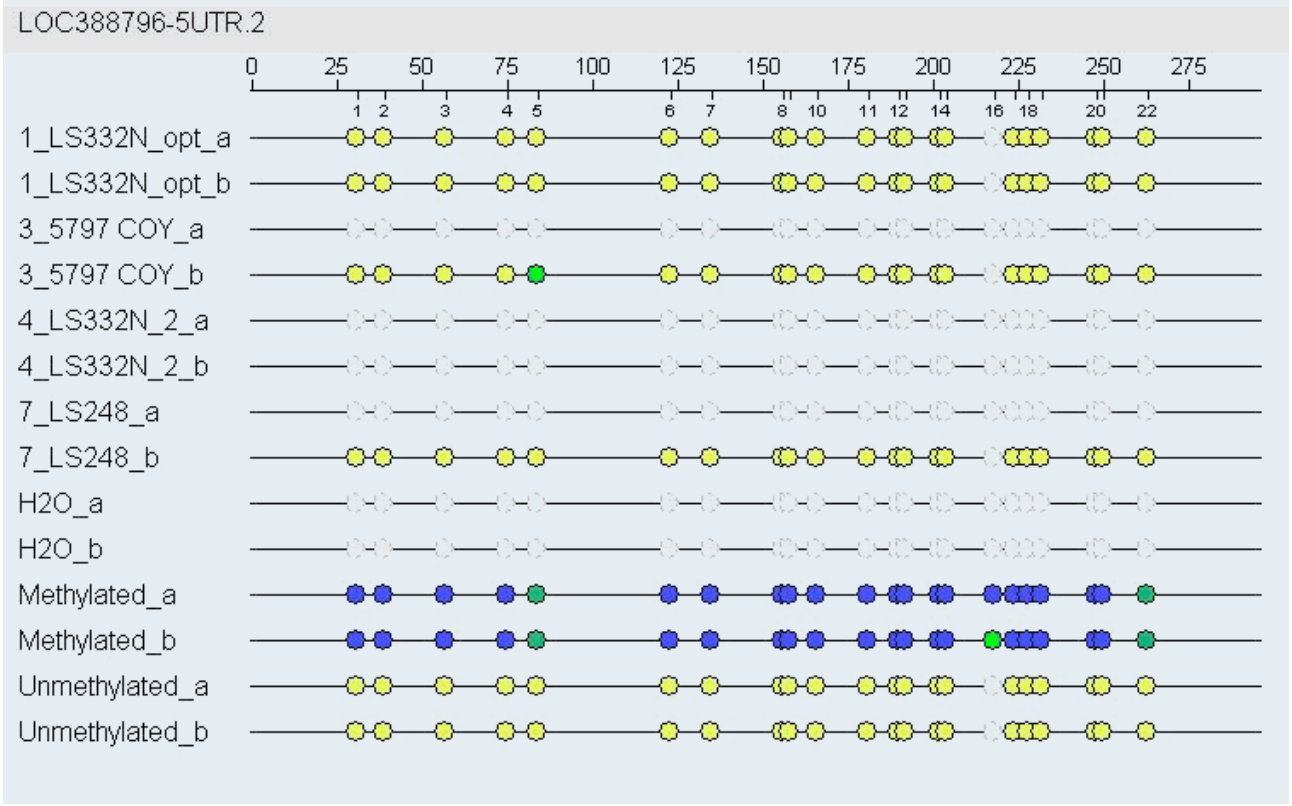
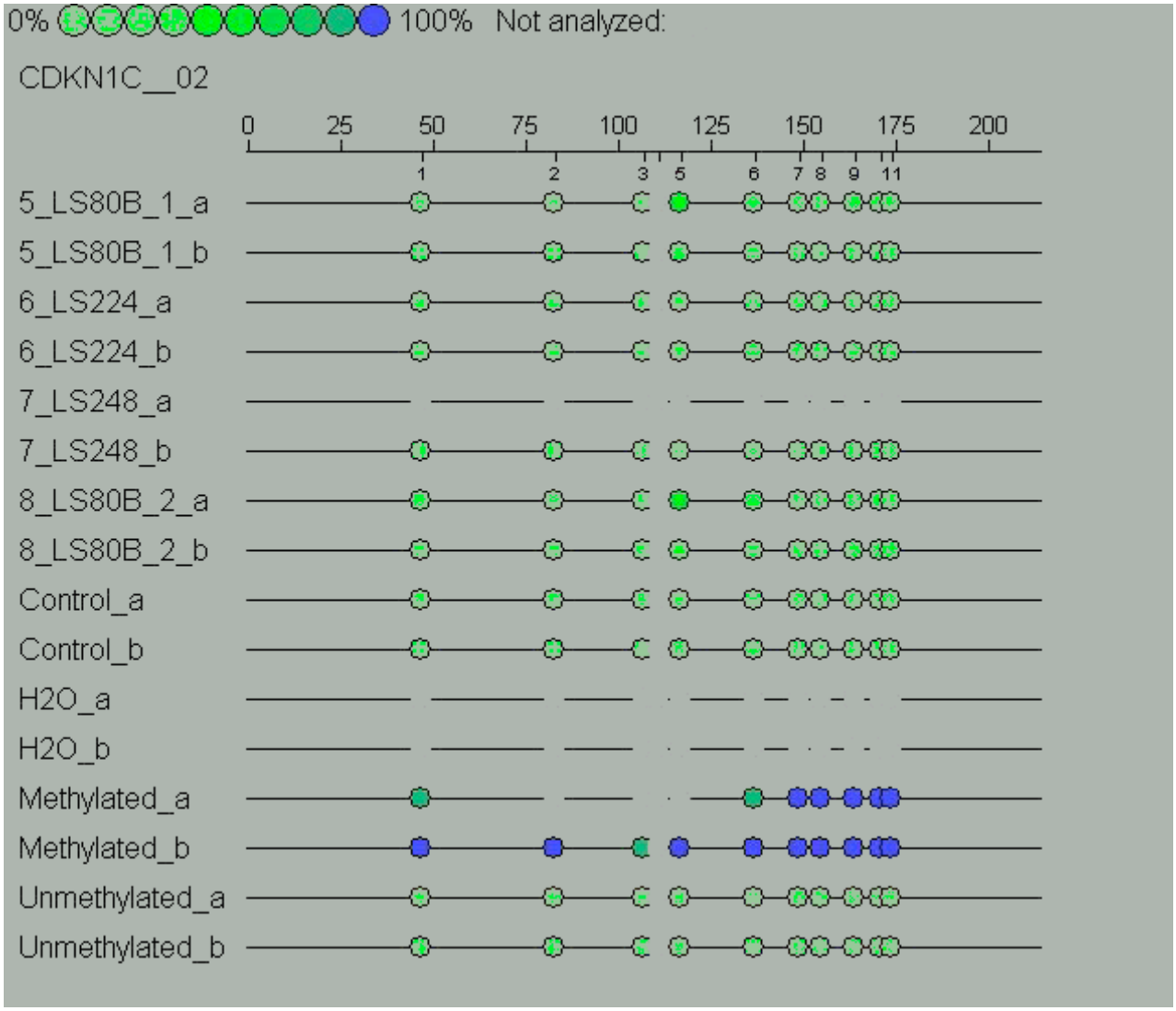
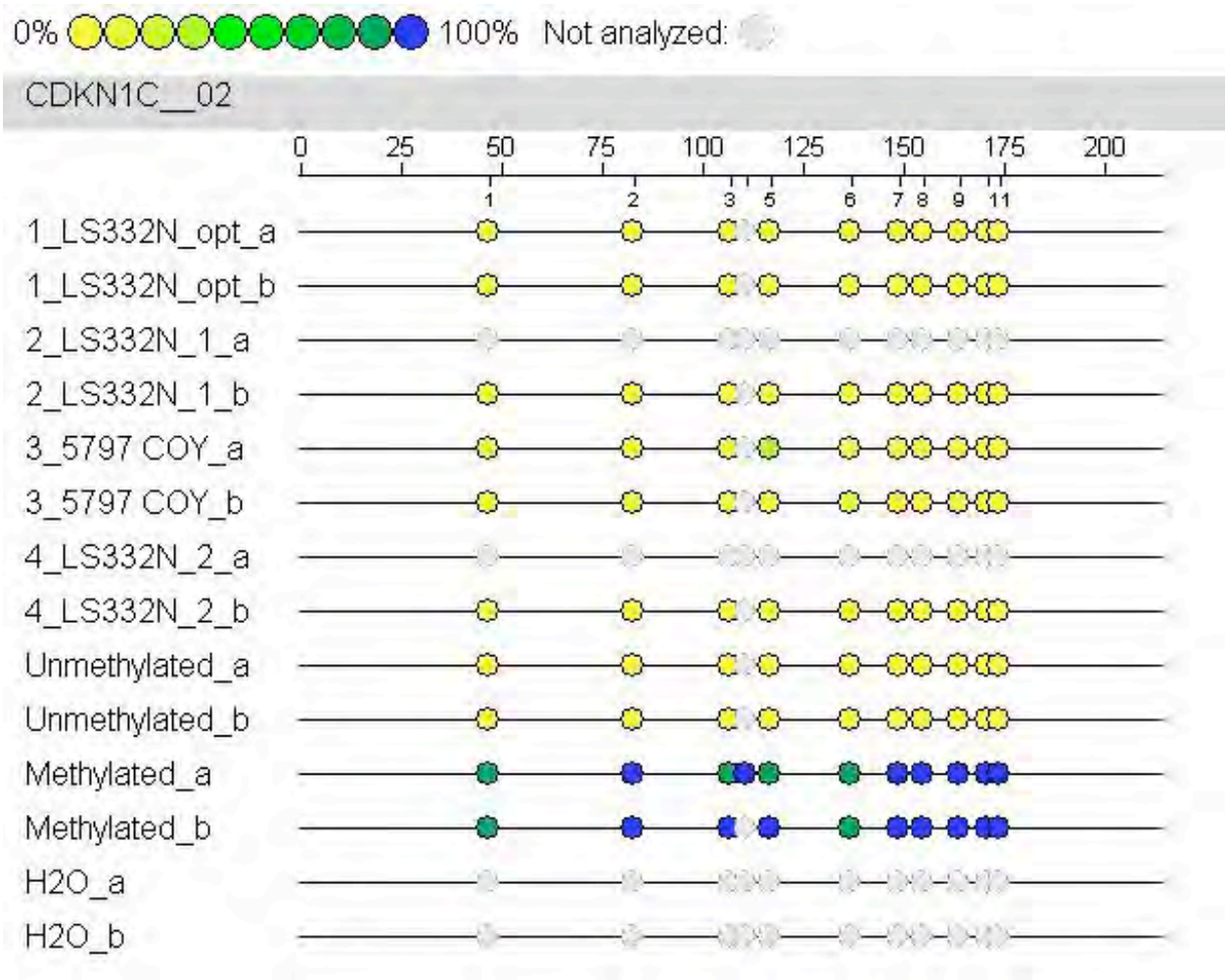
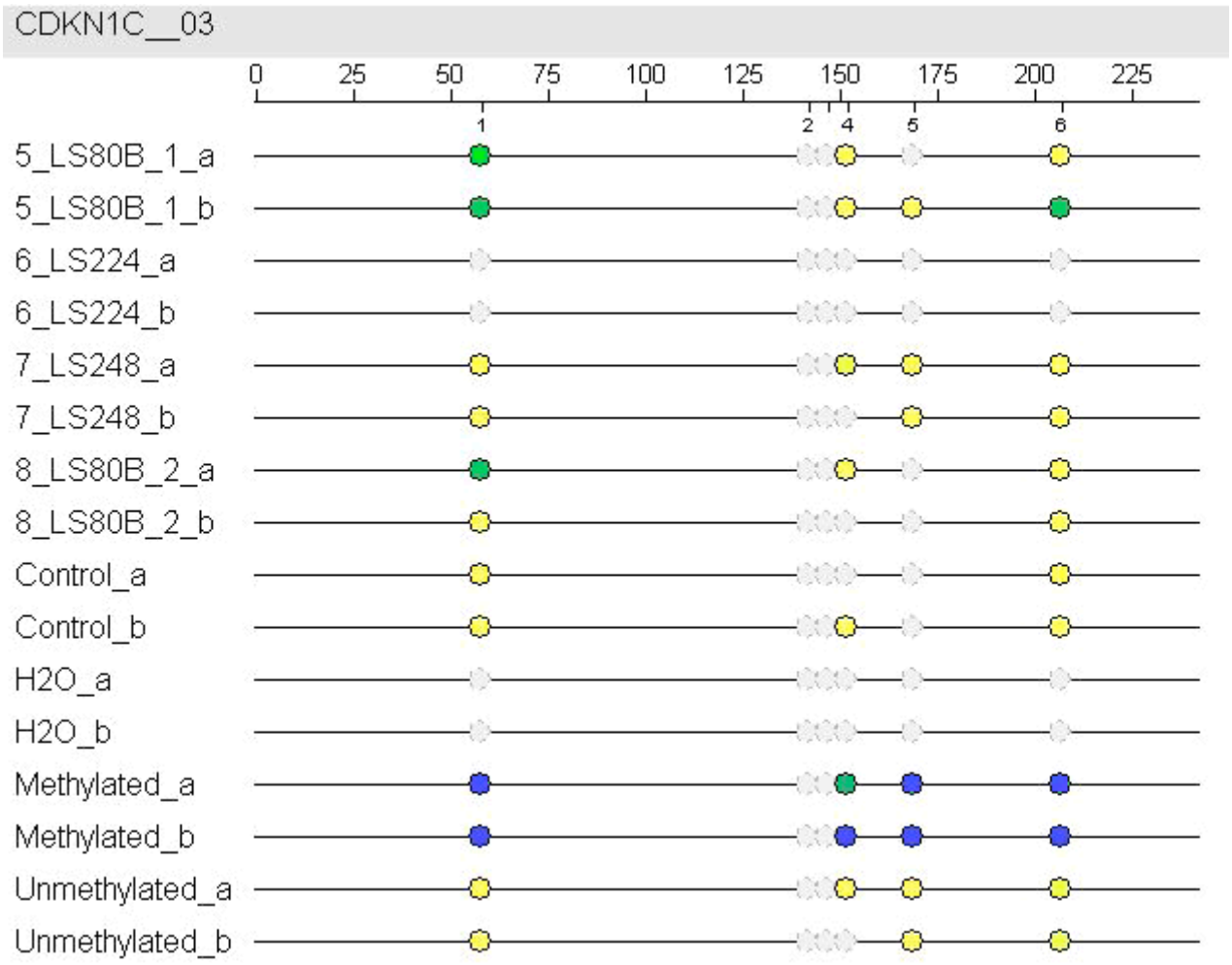

LOC388796-5UTR.2
0 25 50 75 100 125 150 175 200 225 250 275
1 2 3 4 5 6 7 8 10 11 12 14 16 18 20 22
1_LS332N_opt_a
1_LS332N_opt_b
3_5797 COY_a
3_5797 COY_b
4_LS332N_2_a
4_LS332N_2_b
7_LS248_a
7_LS248_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b

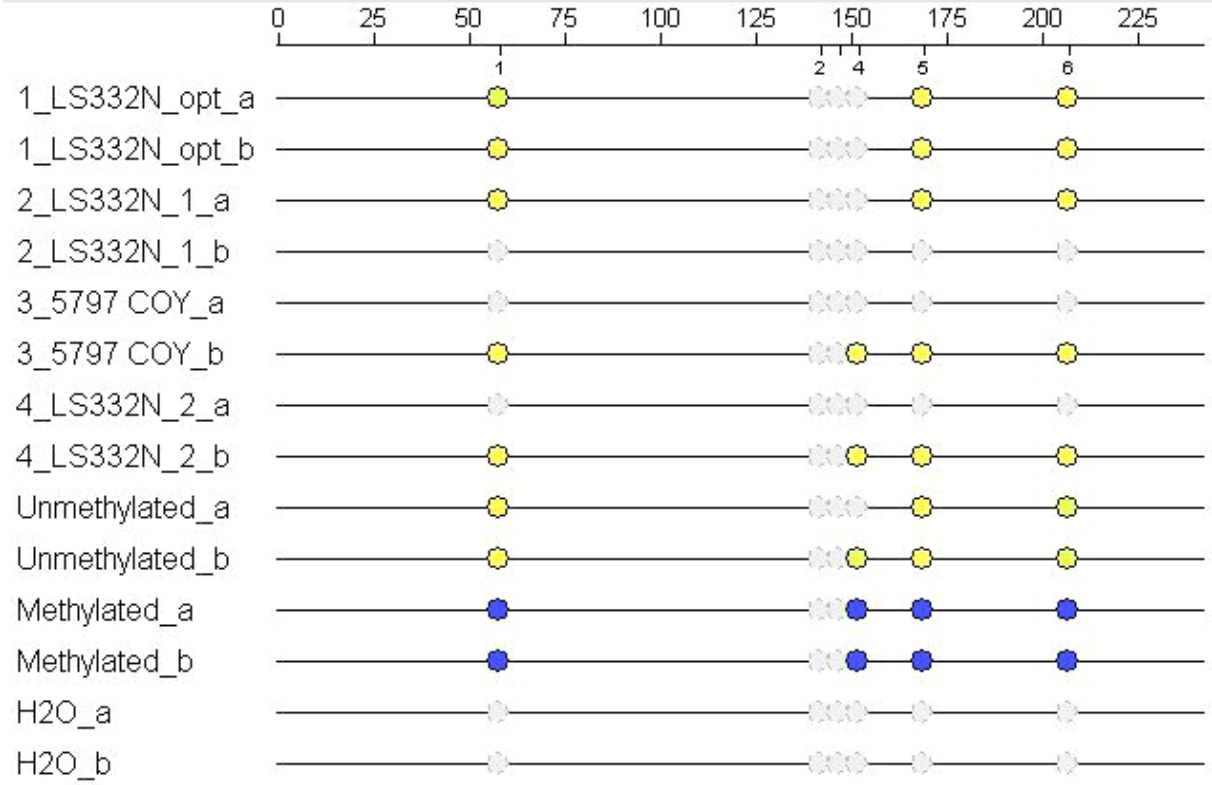
Figure 10: Quantitative DNA methylation analysis in ovarian cancers and normal ovarian epithelium for CTCF sites associated with JAG1, PDGF, THOC3 and the 5'UTR of LOC388796. LCM captured DNA from three ovarian cancers (LS80B, LS224, and LS240) and two normal samples (LS332N and 5797 COY) was treated with sodium bisulfite. Converted DNAs were amplified with primers spanning CTCF binding sites or promoter regions for the indicated genes. PCR products were then subject to basespecific cleavage which depends on the presence of methylated cytosine in the original DNA sample. Cleavage products are then quantitatively analyzed by MALDI-TOF mass spectrometry. MALDI TOF MS analysis of the cleavage product results in a distinct signal pattern from the methylated and non-methylated template DNA. Individual methylation ratios for CpGs within a target sequence were determined and relative methylation ratios assessed in a range between 0%-100% methylation as indicated above each panel as a gradation of color change from yellow (0% methylation) to blue (100% methylation) with a standard deviation of 5%. Amplicon sizes are shown at the top of each data set (for example, JAG1 CTCF site 1 is approximately 225 bp). CpGs interrogated are successively numbered below the amplicon size bar (for example, JAG1 CTCF site 1 contains 7 CpGs of which methylation status was determinable for 5; CpGs 2-6). All amplicons were analyzed in duplicate as indicated (a and b). Control DNA was obtained from SV40 transformed lymphoblasts. Fully methylated DNA (*in vitro* enzymatically methylated genomic DNA) as well as fully unmethylated DNA (chemically and enzymatically treated genomic DNA) were used as controls.


0%
100%
Not analyzed:
CDKN1C__02
0
25
50
75
100
125
150
175
200
1
2
3
5
6
7
8
9
11
5_LS80B_1_a
5_LS80B_1_b
6_LS224_a
6_LS224_b
7_LS248_a
7_LS248_b
8_LS80B_2_a
8_LS80B_2_b
Control_a
Control_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b

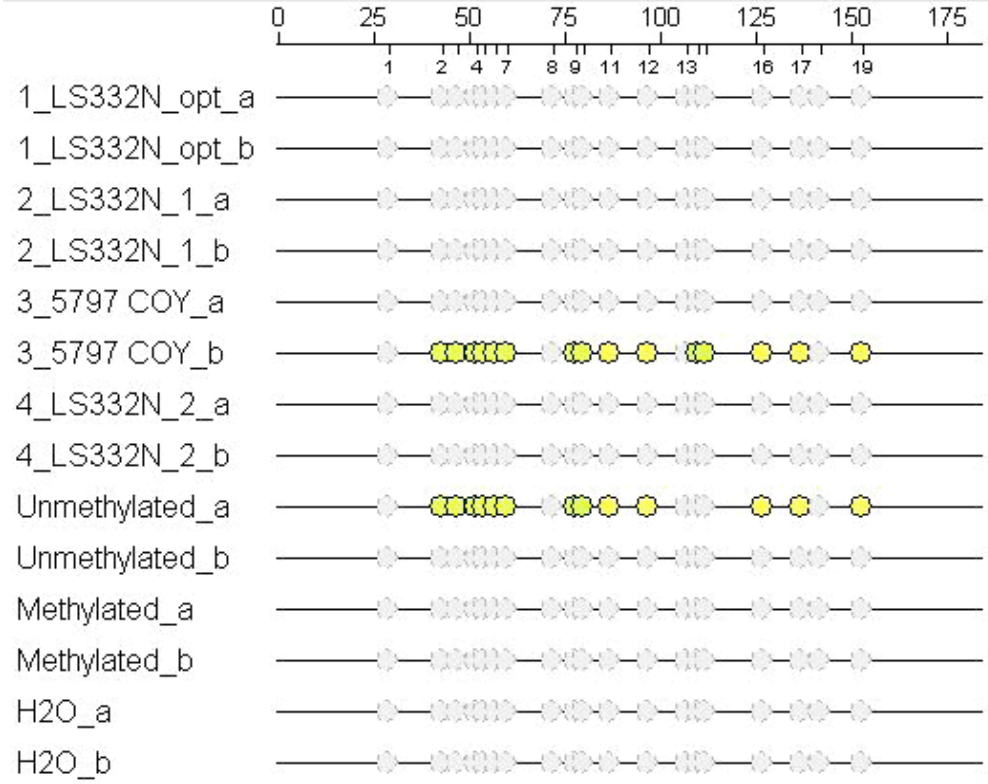

0%
100%
Not analyzed:
CDKN1C__02
0
25
50
75
100
125
150
175
200
1
2
3
5
6
7
8
9
11
1_LS332N_opt_a
1_LS332N_opt_b
2_LS332N_1_a
2_LS332N_1_b
3_5797 COY_a
3_5797 COY_b
4_LS332N_2_a
4_LS332N_2_b
Unmethylated_a
Unmethylated_b
Methylated_a
Methylated_b
H2O_a
H2O_b


CDKN1C__03
0
25
50
75
100
125
150
175
200
225
1
2
4
5
6
5_LS80B_1_a
5_LS80B_1_b
6_LS224_a
6_LS224_b
7_LS248_a
7_LS248_b
8_LS80B_2_a
8_LS80B_2_b
Control_a
Control_b
H2O_a
H2O_b
Methylated_a
Methylated_b
Unmethylated_a
Unmethylated_b

CDKN1C__03



LOC388796_01



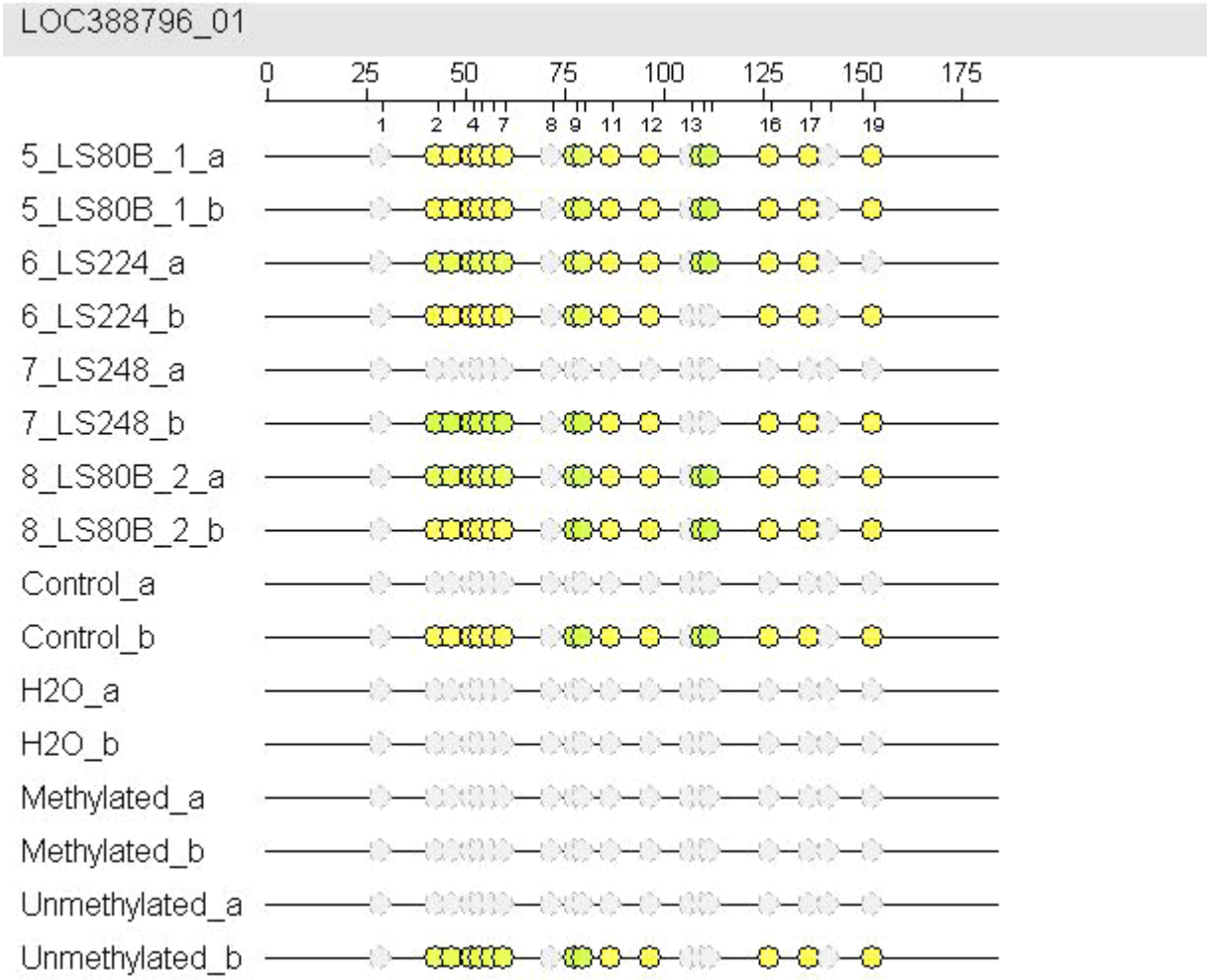
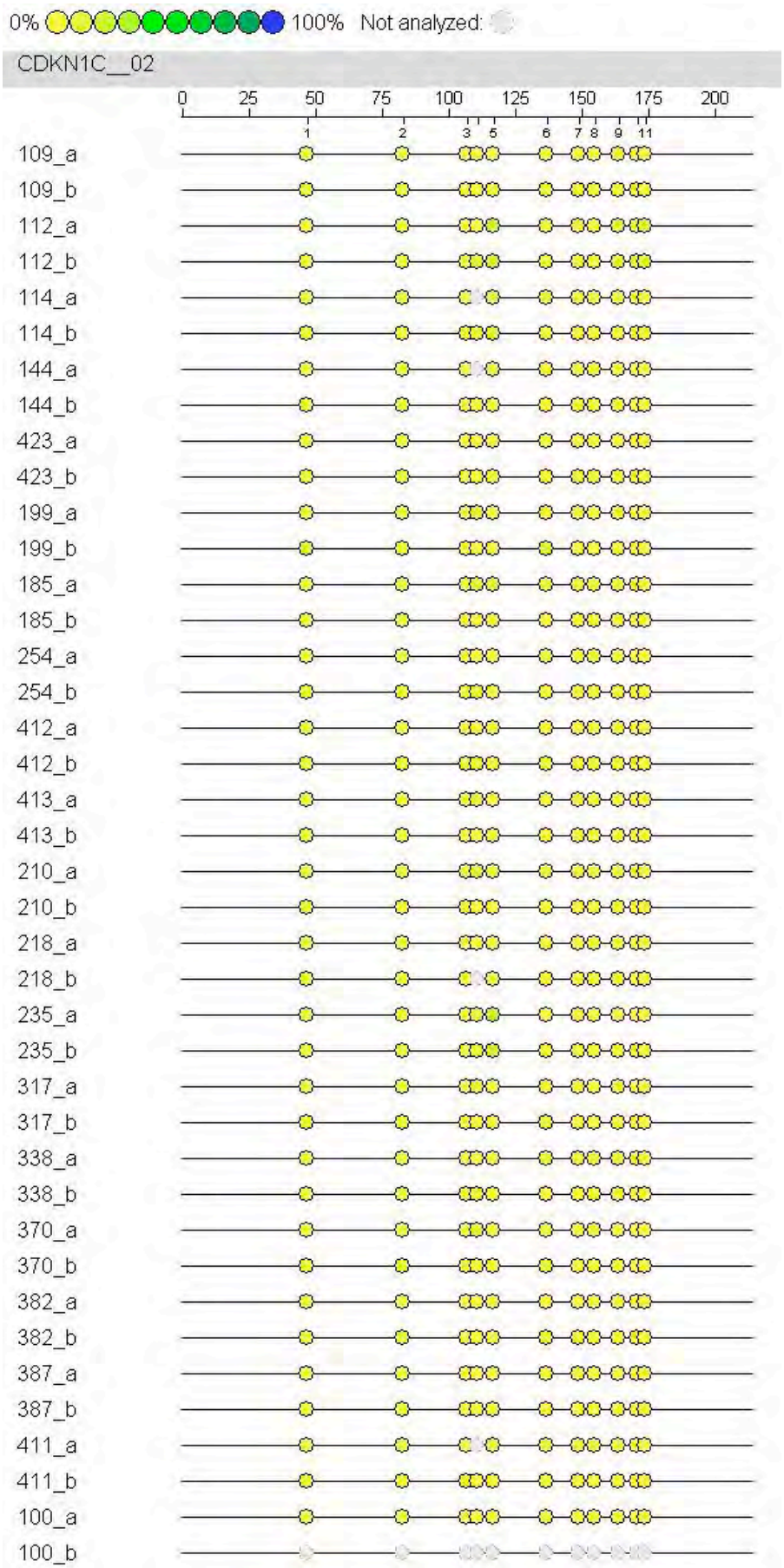
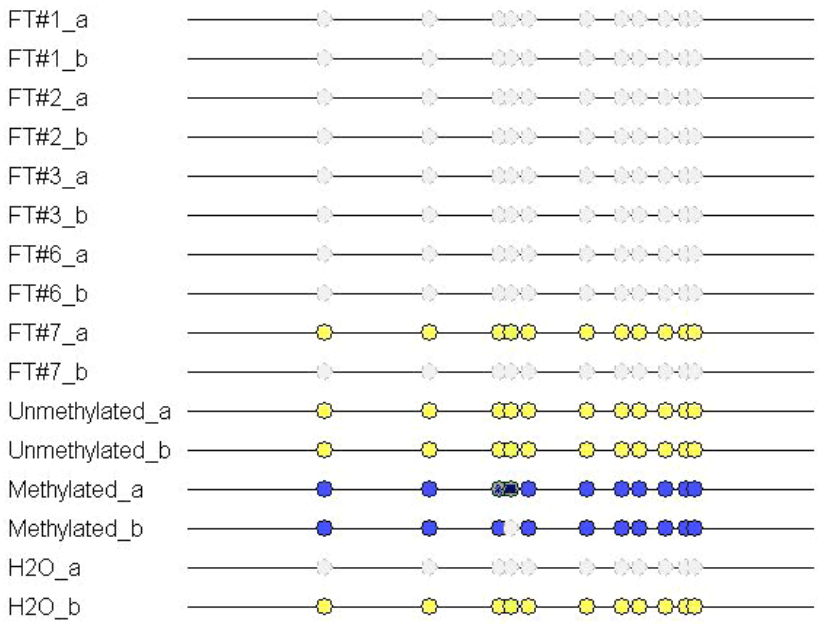
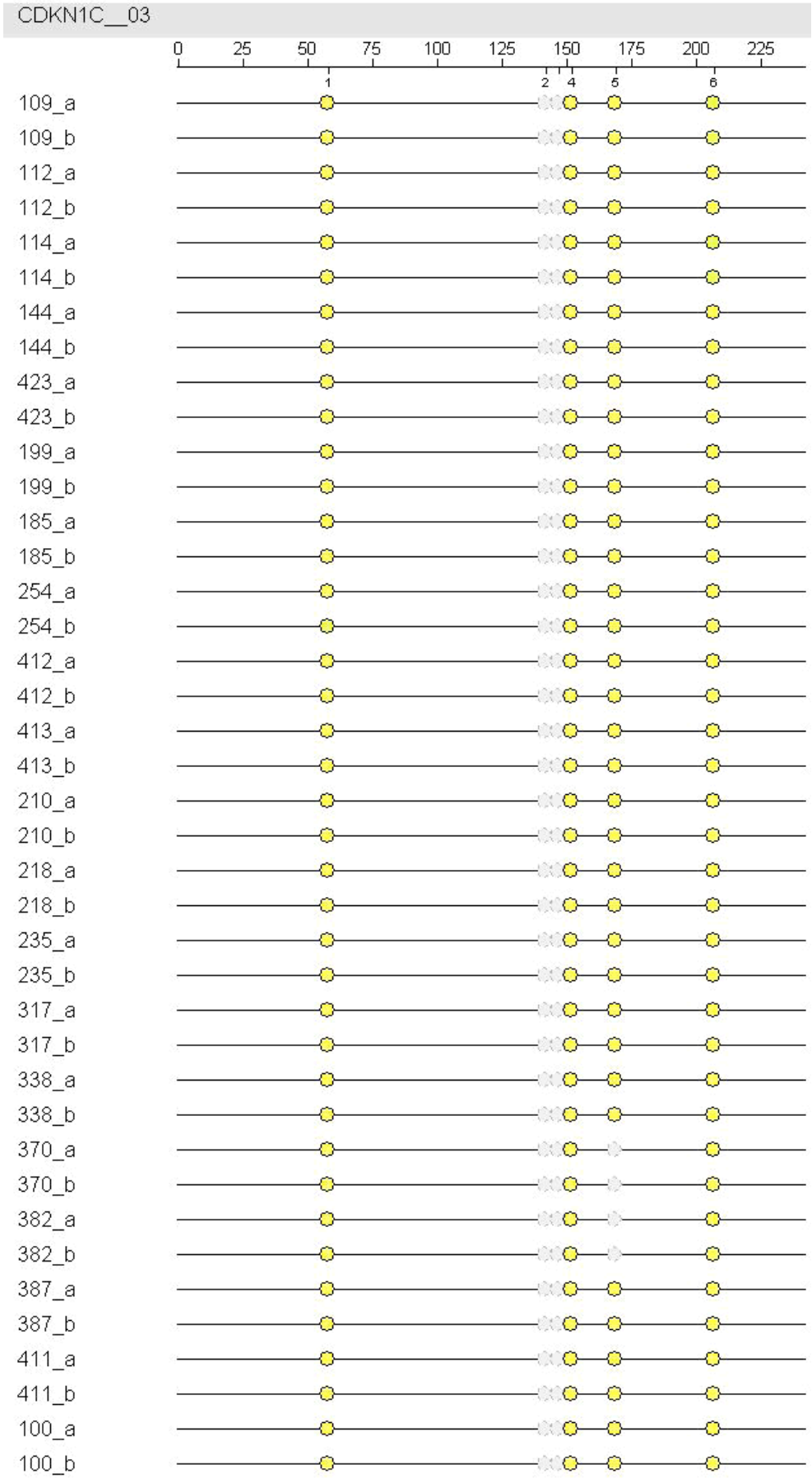


Figure 11: Quantitative DNA methylation analysis in ovarian cancers and normal ovarian epithelium for CTCF sites associated with CDNK1C and the 5'UTR of LOC388796. LCM captured DNA from three ovarian cancers (LS80B, LS224, and LS240) and two normal samples (LS332N and 5797 COY) was treated with sodium bisulfite. Converted DNAs were amplified with primers spanning CTCF binding sites or promoter regions for the indicated genes. PCR products were then subject to basespecific cleavage which depends on the presence of methylated cytosine in the original DNA sample. Cleavage products are then quantitatively analyzed by MALDI-TOF mass spectrometry. MALDI TOF MS analysis of the cleavage product results in a distinct signal pattern from the methylated and non-methylated template DNA. Individual methylation ratios for CpGs within a target sequence were determined and relative methylation ratios assessed in a range between 0%-100% methylation as indicated above each panel as a gradation of color change from yellow (0% methylation) to blue (100% methylation) with a standard deviation of 5%. Amplicon sizes are shown at the top of each data set (for example, CDNK1C CTCF site 2 is approximately 200 bp). CpGs interrogated are successively numbered below the amplicon size bar (for example, CDNK1C CTCF site 2 contains 11 CpGs of which methylation status was determinable for all but CpG 4). All amplicons were analyzed in duplicate as indicated (a and b). Control DNA was obtained from SV40 transformed lymphoblasts. Fully methylated DNA (*in vitro* enzymatically methylated genomic DNA) as well as fully unmethylated DNA (chemically and enzymatically treated genomic DNA) were also used as controls.


0%
100%
Not analyzed:
CDKN1C__02
0
25
50
75
100
125
150
175
200
1
2
3
5
6
7
8
9
11
109_a
109_b
112_a
112_b
114_a
114_b
144_a
144_b
423_a
423_b
199_a
199_b
185_a
185_b
254_a
254_b
412_a
412_b
413_a
413_b
210_a
210_b
218_a
218_b
235_a
235_b
317_a
317_b
338_a
338_b
370_a
370_b
382_a
382_b
387_a
387_b
411_a
411_b
100_a
100_b


FT#1_a
FT#1_b
FT#2_a
FT#2_b
FT#3_a
FT#3_b
FT#6_a
FT#6_b
FT#7_a
FT#7_b
Unmethylated_a
Unmethylated_b
Methylated_a
Methylated_b
H2O_a
H2O_b


CDKN1C__03
0
25
50
75
100
125
150
175
200
225
1
2
4
5
6
109_a
109_b
112_a
112_b
114_a
114_b
144_a
144_b
423_a
423_b
199_a
199_b
185_a
185_b
254_a
254_b
412_a
412_b
413_a
413_b
210_a
210_b
218_a
218_b
235_a
235_b
317_a
317_b
338_a
338_b
370_a
370_b
382_a
382_b
387_a
387_b
411_a
411_b
100_a
100_b

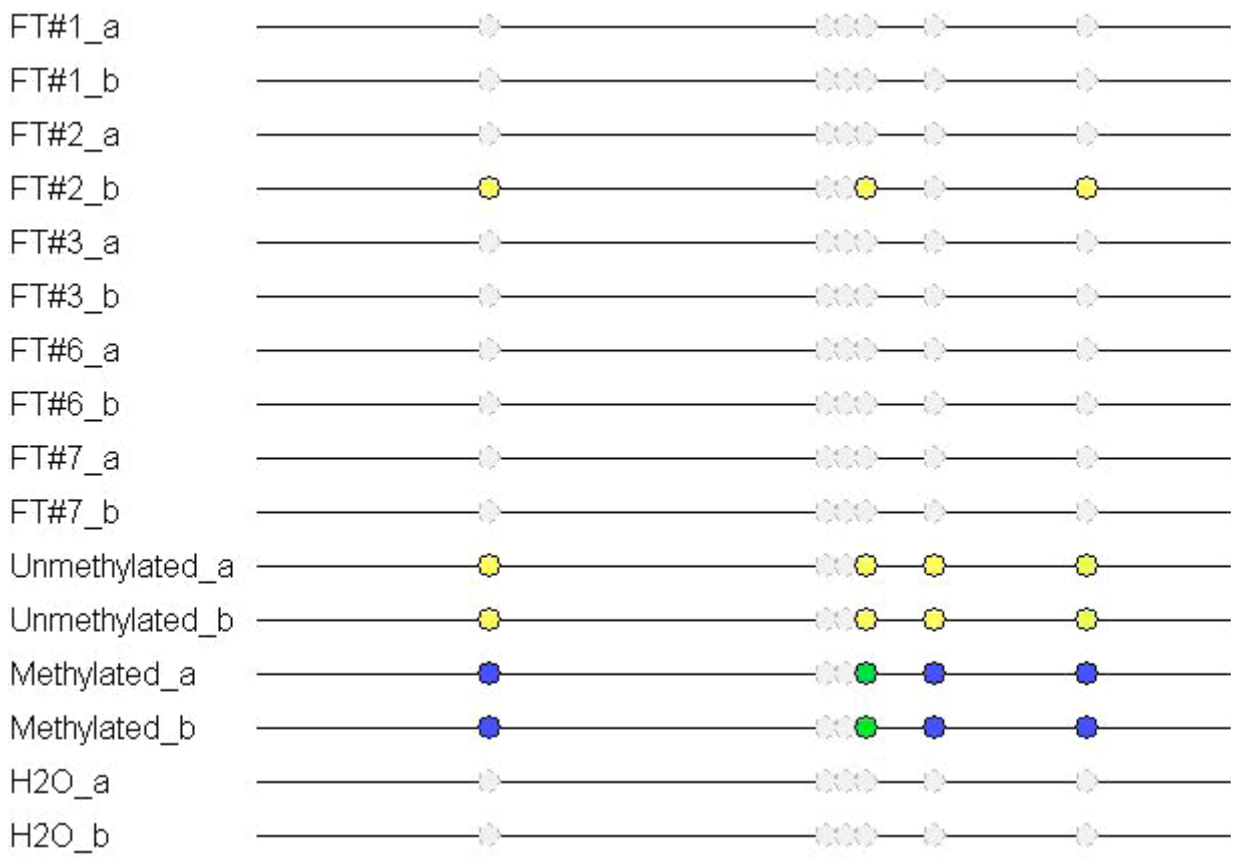


Figure 12. Quantitative DNA methylation analysis in ovarian cancers and normal ovarian FT epithelium for the CTCF site associated with CDNK1C_02 and _03 in 20 ovarian cancers and 5 normal FT epithelial lines.

NEOPLASIA
www.neoplasia.com

Volume 12 Number 12 December 2010 pp. 1–10 1

Identification of a Preneoplastic Gene Expression Profile in Tubal Epithelium of BRCA1 Mutation Carriers[1,2,3]

Joshua Z. Press*, Kaitlyn Wurz*,†, Barbara M. Norquist*, Ming K. Lee†, Christopher Pennil†, Rochelle Garcia‡, Piri Welcsh†, Barbara A. Goff* and Elizabeth M. Swisher*,†

*Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, University of Washington School of Medicine, Seattle, WA, USA; †Division of Medical Genetics, Department of Medicine, University of Washington School of Medicine, Seattle, WA, USA; ‡Department of Pathology, University of Washington School of Medicine, Seattle, WA, USA

Abstract

Microinvasive carcinomas and high-grade intraepithelial neoplasms are commonly discovered within the fallopian tube of BRCA1 mutation carriers at the time of risk-reducing salpingo-oophorectomy, suggesting that many BRCA1-mutated ovarian carcinomas originate in tubal epithelium. We hypothesized that changes in gene expression profiles within the histologically normal fallopian tube epithelium of *BRCA1* mutation carriers would overlap with the expression profiles in *BRCA1*-mutated ovarian carcinomas and represent a *BRCA1* preneoplastic signature. Laser capture microdissection of frozen sections was used to isolate neoplastic cells or histologically normal fallopian tube epithelium, and expression profiles were generated on Affymetrix U133 Plus 2.0 gene expression arrays. Normal-risk controls were 11 women wild type for *BRCA1* and *BRCA2* (WT-FT). WT-FT were compared with histologically normal fallopian tube epithelium from seven women with deleterious *BRCA1* mutations who had foci of at least intraepithelial neoplasm within their fallopian tube (B1-FTocc). WT-FT samples were also compared with 12 *BRCA1* ovarian carcinomas (B1-CA). The comparison of WT-FT *versus* B1-FTocc resulted in 152 differentially expressed probe sets, and the comparison of WT-FT *versus* B1-CA resulted in 4079 differentially expressed probe sets. The BRCA1 preneoplastic signature was composed of the overlap between these two lists, which included 41 concordant probe sets. Genes in the *BRCA1* preneoplastic signature included several known tumor suppressor genes such as *CDKN1C* and *EFEMP1* and several thought to be important in invasion and metastasis such as *E2F3*. The expression of a subset of genes was validated with quantitative reverse transcription–polymerase chain reaction and immunohistochemistry.

Neoplasia (2010) 12, 1–10

Introduction

Ovarian carcinoma is the leading cause of death from gynecologic malignant neoplasms in the developed world. Identification of the early molecular events leading to ovarian carcinoma has been hindered by the lack of an identifiable preneoplastic lesion and the limited occurrence of early-stage neoplasms. Although it has been proposed that ovarian carcinoma originates from the surface epithelium of the ovary and/or the epithelial lining of ovarian inclusion cysts, there have been few reports of intraepithelial neoplasms at these sites [1,2].

Address all correspondence to: Joshua Z. Press, MD, Department of Obstetrics and Gynecology, University of Washington Medical Center, Box 356460, Seattle, WA 98195. E-mail: jzp@u.washington.edu

[1]This study was supported by the Yvonne Betson Trust, DOD OC073389 (to E.M.S.) and 1R01CA131965 (E.M.S.). No authors have conflicts of interest to declare.
[2]This study was presented in abstract form at the 40th Annual Society of Gynecologic Oncologists meeting in San Antonio, TX, 2009.
[3]This article refers to supplementary materials, which are designated by Tables W1 and W2 and Figures W1 to W4 and are available online at www.neoplasia.com.
Received 22 July 2010; Revised 6 September 2010; Accepted 9 September 2010

DOI 10.1593/neo.101044

Alternatively, there has been increasing evidence that many ovarian carcinomas originate within the fallopian tube [3]. Fallopian tube epithelium can exhibit areas of increased proliferation and cytologic atypia, called intraepithelial neoplasia (IEN). Most ovarian carcinomas are of serous histology and frequently exhibit mutations in the critical cell cycle regulator p53 [4]. Severe IEN in fallopian tubes has been found in conjunction with mullerian malignant neoplasms, particularly serous carcinomas of ovarian, uterine, or peritoneal origin [5,6]. Identical p53 mutations have been identified in tubal IEN and coexisting sporadic serous carcinoma [7], suggesting that genetic disruption within the fallopian tube may progress to ovarian carcinoma.

Further evidence for a tubal origin is suggested by the high prevalence of occult fallopian tube carcinomas identified among *BRCA1* and *BRCA2* mutation carriers undergoing risk-reducing salpingo-oophorectomy (RRSO). Although the lifetime risk of ovarian carcinoma in the general population is only 1% to 2%, women who inherit mutations in the *BRCA1* and *BRCA2* genes have up to a 50% lifetime risk of ovarian carcinoma [8]. These high-risk women are frequently discovered to have occult neoplasms at the time of RRSO, and 57% to 100% of these lesions arise in the fallopian tube [9–11]. Fallopian tube epithelium frequently contains areas that have been termed *p53 foci* (also referred to as p53 signatures), which overexpress p53 and have increased expression of the proliferation marker Ki-67 [12]. These tubal p53 foci are more frequent in tubes from BRCA1 and BRCA2 mutation carriers compared with normal-risk women, and they have also been shown to exhibit decreased expression of the tumor suppressor protein p27 [13]. These observations have resulted in the proposal of a new paradigm for ovarian carcinoma, in which the fallopian tube epithelium acquires a sequence of molecular abnormalities leading to an *in situ* or invasive neoplasm, which exfoliates and spreads to the ovary and peritoneum [3]. Validating the role of the fallopian tube in ovarian carcinoma carcinogenesis will require additional studies, such as comparative analysis of gene expression between wild-type and high-risk fallopian tubes.

We obtained frozen fallopian tube tissue from seven women with *BRCA1* mutations found to have occult invasive carcinomas or severe IEN in the fallopian tube on final pathologic examination. We hypothesized that the histologically normal tubal epithelium from these women would possess a gene expression profile that would reflect early alterations in gene expression contributing to the development of carcinoma. By comparing the gene expression profiles between these high-risk fallopian tubes and histologically normal fallopian tubes from women with wild-type *BRCA1* and *BRCA2*, we identified a set of genes potentially important in the development of *BRCA1*-associated carcinomas. We hypothesized that genes important in *BRCA1* ovarian carcinogenesis would have similarly altered expression patterns in *BRCA1* carcinomas. Therefore, we used the expression patterns in *BRCA1* ovarian carcinomas to further define the genes of interest in *BRCA1* tubal epithelium.

Materials and Methods

Study Design and Sample Selection

All tissues and clinical information were obtained from the University of Washington Gynecologic Oncology Tissue Bank according to an institutional review board–approved protocol. To maximize the likelihood of identifying biologically important gene differentially expressed between histologically normal BRCA wild-type fallopian tubes and high-risk fallopian tubes from *BRCA1* mutation carriers, we specifically selected *BRCA1* mutation carriers possessing occult microinvasive or high-grade intraepithelial fallopian tube neoplasm to create the gene profile (B1-FTocc). In addition, to minimize the false discovery rate, we also identified genes differentially expressed between the BRCA wild-type fallopian tube epithelium (WT-FT) and invasive *BRCA1* carcinomas (B1-CA). We limited our *BRCA1* preneoplastic profile to genes showing concordant up-regulation or down-regulation in both B1-FTocc and B1-CA. Thirty patients were analyzed to create the *BRCA1* preneoplastic gene signature: 11 histologically normal fallopian tube epithelium from women with wild-type *BRCA1* and *BRCA2* (WT-FT), 7 histologically normal fallopian tube epithelium from women with deleterious *BRCA1* mutations and documented occult microinvasive or high-grade intraepithelial fallopian tube carcinoma (B1-FTocc), and 12 high-grade serous ovarian carcinomas from women with deleterious *BRCA1* mutations (B1-CA). The characteristics of these patients are shown in Table 1. We chose WT-FT samples to match the age and menopausal distribution of the B1-FTocc cases. Some women in the WT-FT group had a personal history of breast cancer or a family history of breast cancer; however, women were excluded from the WT-FT group if they had a family history of ovarian cancer. All WT-FT control women had had full gene sequencing by Myriad genetics, and those who did not have comprehensive rearrangement testing performed by Myriad were screened with Multiplex Ligation–dependent Probe Amplification (MRC-Holland BV, Amsterdam, Holland) according to the manufacturer's instructions in our laboratory using normal DNA extracted from lymphocytes. For the B1-FTocc samples, the histologically normal epithelium was obtained from the same fallopian tube discovered to contain the occult fallopian tube neoplasm. Three of the seven B1-FTocc women were premenopausal at the time of oophorectomy.

Laser Capture Microdissection, RNA Amplification, and Gene Expression Chips

Tissues samples had been collected at the time of RRSO or ovarian carcinoma cytoreductive surgery and were immediately frozen in the operating room in liquid nitrogen in Tissue-Tek OCT (Alphen aan den Rijn, the Netherlands). For RRSO specimens, a small piece of tubal fimbriae was collected for the tissue bank. A frozen section of that tissue was stained with hematoxylin and eosin to confirm normal histologic diagnosis and rule out neoplasia in the research specimen. The remaining fallopian tube tissues from these cases were then subjected to serial sectioning by the pathologist to look for intraepithelial carcinoma or invasive carcinoma. All stored samples were subjected to the identical protocol of laser capture microdissection (LCM), linear RNA amplification, and microarray production. Hematoxylin and eosin slides from the frozen tissue OCT blocks were reviewed to select blocks with adequate distal fimbriated fallopian tube epithelium. Before LCM, 7-μm frozen sections were cut, adhered onto glass membrane slides (Arcturus, Mountain View, CA), and immediately stored on dry ice. Before LCM, the slides were dehydrated and stained with hematoxylin with the Histogene LCM Frozen Section Staining Kit (Arcturus). Slides were immediately transferred to the Veritas Laser Capture Microdissection system (Arcturus). Fallopian tube epithelium from the distal fimbriated fallopian tube was selectively captured for the fallopian tube samples, and ovarian carcinoma cells were selectively captured for neoplastic samples (Figure W1). Total RNA was isolated, and contaminating DNA was removed using the PicoPure RNA Isolation Kit (Arcturus) as per the company's protocol. The MessageAmp II aRNA amplification Kit (Ambion, Austin, TX) was used to amplify

Table 1. Cases Used to Generate the *BRCA1* Preneoplastic Gene Signature.

Case Identifier	Age (years)	Menopausal Status	*BRCA1/2* Status*	Other Characteristics
WT-FT no. 1	46	Pre	Negative	Personal history of breast cancer
WT-FT no. 2	47	Post	Negative	Personal history of breast cancer
WT-FT no. 3	48	Post	Negative	Personal history of breast cancer
WT-FT no. 4	48	Pre	Negative	No personal history of cancer
WT-FT no. 5	49	Post	Negative	Personal history of breast cancer
WT-FT no. 6	50	Pre	Negative	Personal history of breast cancer
WT-FT no. 7	52	Pre	Negative	Personal history of breast cancer
WT-FT no. 8	54	Post	Negative	Personal history of breast cancer
WT-FT no. 9	55	Post	Negative	Personal history of breast DCIS
WT-FT no. 10	61	Post	Negative	No personal history of cancer
WT-FT no. 11	61	Post	Negative	Personal history of breast cancer
B1-FTocc no. 1	39	Pre	B1.3109insAA	Microinvasion left fallopian tube
B1-FTocc no. 2	40	Post	B1.M1V (120A>G)	Microinvasion left fallopian tube
B1-FTocc no. 3	47	Pre	B1.2800delAA	High-grade intraepithelial
B1-FTocc no. 4	49	Pre	B1.3795del4	Microinvasion right fallopian tube†
B1-FTocc no. 5	53	Post	B1.del ex14-20	High-grade intraepithelial
B1-FTocc no. 6	62	Post	B1.C61G	High-grade intraepithelial
B1-FTocc no. 7	63	Post	B1.2800delAA	Microinvasion left fallopian tube
B1-CA no. 1	40	Pre	B1.2576.delC	Stage IIIC, grade 3, serous carcinoma
B1-CA no. 2	41	Pre	B1.185delAG	Stage IIIC, grade 3, serous carcinoma
B1-CA no. 3	44	Pre	B1.2798del4	Stage IIIC, grade 3, serous carcinoma
B1-CA no. 4	49	Pre	B1.3795del4	Stage IIIC, grade 3, serous carcinoma†
B1-CA no. 5	50	Post	B1.5382insC	Stage IA, grade 3, undifferentiated
B1-CA no. 6	51	Post	B1.3171ins5	Stage IIIC, grade 3, serous carcinoma
B1-CA no. 7	54	Post	B1.2594delC	Stage IV, grade 3, serous carcinoma
B1-CA no. 8	54	Post	B1.del_exon14	Stage IIIC, grade 3, serous carcinoma
B1-CA no. 9	55	Post	B1.M1V (120A>G)	Stage IIC, grade 3, serous carcinoma
B1-CA no. 10	57	Post	B1.5382insC	Stage IIIC, grade 3, mixed serous/endo
B1-CA no. 11	57	Post	B1.5382insC	Stage IIIC, grade 3, serous carcinoma
B1-CA no. 12	65	Post	B1.5382insC	Stage IIIC, grade 3, serous carcinoma

*Negative cases were wild-type by full sequencing as well as by comprehensive testing for gene rearrangements. WT-FT indicates histologically normal fallopian tube epithelium from *BRCA1* wild-type patients; B1-FTocc, histologically normal fallopian tube epithelium from patients with deleterious *BRCA1* mutations and at least high-grade intraepithelial fallopian tube neoplasm; B1-CA, tumor tissue from patients with deleterious *BRCA1* mutations.

†B1-FTocc no. 4 and B1-CA no. 4 are from the same individual who had both microscopic invasive neoplasm within the fallopian tube and peritoneal metastasis. DCIS indicates ductal carcinoma in situ. Menstrual phase of WT-FT cases: WT-FT no. 1 and WT-FT no. 7 had proliferative endometrium and WT-FT no. 4 and WT-FT no. 6 did not have hysterectomy performed.

the total RNA once. The quality of each amplified RNA sample was confirmed using Agilent 2000 Bioanalyzer RNA 6000 Pico LabChip Kit (Agilent Technologies, Inc, Santa Clara, CA), and quantity was measured using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE). All labeling, hybridization, and scanning were performed at the University of Washington Centre for Array Technology core facility. Amplified complementary DNA (cDNA) was purified, enzymatically fragmented, and labeled with biotin. Quality and quantity of the purified labeled cDNA product were confirmed before hybridizing to Affymetrix GeneChip U133A Plus 2.0 Arrays (Affymetrix, Inc, Santa Clara, CA). To minimize batch effect, several samples from each study group were included in each batch of array runs.

Development of the Gene Expression Profile

GeneSifter (Seattle, WA) software was used for pairwise gene expression analysis and clustering analysis (Manhattan, Complete Linkage). For pairwise gene expression analysis, a Welch *t* test was used when generating *P* values. To develop a potential *BRCA1* preneoplastic gene expression profile, two independent comparisons were made. First, the 11 WT-FT were compared with the seven B1-FTocc samples, and probe sets were selected, which showed a 1.8-fold differential expression at $P < .01$. To minimize the false discovery rate, we also performed a comparison between the 11 WT-FT and 12 B1-CA and selected probe sets, which showed a 1.8-fold differential expression at $P < .01$. Probe sets were only selected for the *BRCA1* preneoplastic gene expression profile if they demonstrated concordant up-regulation or down-regulation in both the B1-FTocc and the B1-CA. The 1.8-fold was the cutoff at which we had most overlapping genes while still optimizing the ratio of overlapping genes to nonoverlapping genes.

Real-time Quantitative Reverse Transcription–Polymerase Chain Reaction Analysis

Real-time quantitative reverse transcription–polymerase chain reaction (RT-PCR) was used to validate the Affymetrix array results for four genes from the *BRCA1* preneoplastic signature. For each group (WT-FT, B1-FTocc, and B1-CA), five representative cases were selected and analyzed for expression of the four genes (*EFEMP1*, *p57*, *CYP3A5*, and *CSPG5*). TaqMan Gene Expression Assays (Applied Biosystems, Carlsbad, CA) were used for *EFEMP1* (Hs01013942_m1), *p57* (Hs00908986_g1), *CYP3A5* (Hs02511768_s1), and *CSPG5* (Hs00962721_m1), and *GAPDH* was used as the reference gene. All samples were run in triplicate, and the comparative C_t method was used for relative quantitation using ABI PRISM Sequence Detection Software (Applied Biosystems). Target gene C_t values were normalized to *GAPDH*.

Interrogation of the BRCA1 Preneoplastic Gene Signature Using Independent Samples

Additional fallopian tube samples that were not used to create the gene expression signature were selected to interrogate the *BRCA1* preneoplastic gene signature. To test the intrasample reproducibility of the tubal expression profiles, three duplicates of the B1-FTocc samples were analyzed (Table W1). For two of these B1-FTocc – DUP cases (B1-FTocc no. 2 – DUP and BT-FTocc no. 6 – DUP), expression arrays were created using tissue from the fallopian tube contralateral

to the microinvasive or high-grade intraepithelial lesion. For the duplicate of B1-FTocc no. 1, LCM was performed a second time using separate sections obtained approximately 100 μm further into the frozen tissue block. In addition, 12 *BRCA1*-mutated fallopian tubes from postmenopausal women, which did not contain occult lesions (B1-FT), were also analyzed (Table W2). The expression profiles were generated by comparing to the same set of WT-FT cases. These additional expression profiles were then analyzed by combining them one at a time with the original 30 samples that had been used to create the premaligant gene signature. Unsupervised hierarchical clustering (Manhattan, Complete Linkage) using the probe sets from the *BRCA1* preneoplastic gene signature was performed for each combination to determine whether the additional cases contained the *BRCA1* preneoplastic expression profile and would therefore cluster with the B1-FTocc cases. In addition, for 10 of the B1-CA carcinomas with adequate DNA available, DNA was sequenced for p53 exons 4 to 10 as previously described [14].

Ki-67 Immunohistochemistry

To validate the array gene expression data for *MKI67* (antigen identified by monoclonal antibody Ki-67), we performed immunohistochemistry on a larger set of fallopian tube samples, which had been stored as paraffin blocks. Most of these cases had right and left distal fallopian tube tissues available, and an average Ki-67 staining score was obtained from both tubes. Fallopian tube tissues from 26 *BRCA1* wild-type cases were compared with fallopian tube tissues from 52 *BRCA1* mutation carriers without carcinoma obtained at RRSO. Paraffin sections were deparaffinized and sequentially rehydrated, and endogenous peroxidases were blocked. Heat-mediated antigen retrieval was performed in a citrate buffer (Antigen Unmasking Solution; Vector Laboratories, Burlingame, CA). Slides were incubated with the Ki-67 mouse monoclonal antibody MIB-1, diluted 1:100 (Dako, Copenhagen, Denmark). Sections were washed with phosphate-buffered saline and incubated with secondary antibody (horseradish peroxidase–antimouse; Vector Laboratories). DAB was used to visualize antibody complexes, and sections were counterstained with hematoxylin. Negative and positive controls were assessed for each run. Slides were scored by two independent observers blinded to case designation. The percentages of positive epithelial cells were scored (0 = none, 1 = 1%, 2 = 2%-4%, 3 = 5%-15%, 4 ≥15%). A Mann-Whitney test was used to compare the staining results.

Results

Affymetrix Gene Expression

There were 18,600 probe sets expressed on the Affymetrix chips, which showed quality more than 0.7 in all samples. There were 152 probe sets with significant differential expression (>1.8-fold, $P < .01$) between the WT-FT and B1-FTocc. There were 4079 probe sets with significant differential expression (>1.8-fold, $P < .01$) between the WT-FT and B1-CA. The 152 probe sets differentially expressed from the BI-FTocc were compared with the 4079 differentially expressed probe sets in the B1-CA (Figure 1). The overlap between the two differentially expressed probe sets consisted of 29 probe sets downregulated in both groups, 12 probe sets upregulated in both groups, and 7 probe sets showing contradictory expression (up-regulation in one comparison and downregulated in the other). The 41 probe sets demonstrating concordant up-regulation or down-regulation in both comparisons comprised the *BRCA1* preneoplastic gene signature and are shown in Table 2 and Figure 2.

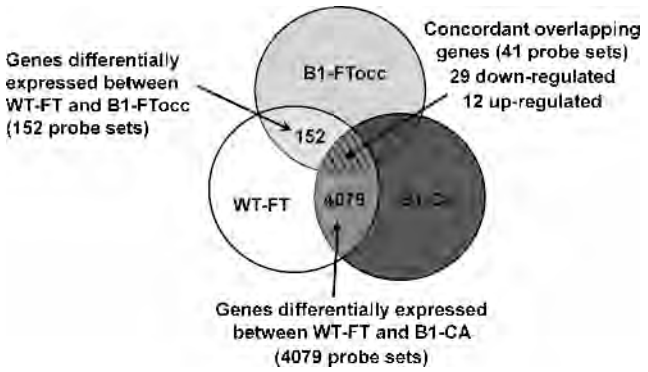


Figure 1. Diagram illustrating the protocol used to define the BRCA1 preneoplastic gene signature: Pairwise comparison between the WT-FT group and the B1-FTocc group (fold change ≥ 1.8, $P < .01$) identified 152 differentially expressed probe sets. Pairwise comparison between the WT-FT group and the B1-CA group (fold change ≥ 1.8, $P < .01$) identified 4079 differentially expressed probe sets. To minimize the false discovery rate, probe sets were only included in the gene signature with concordant down-regulation or up-regulation in both pairwise comparisons (hatched region).

To test the significance of the overlap in differentially expressed genes, we created a simulation in which there were 4079 randomly selected expressed clones in one group and 152 in a second group from 18,600 expressed clones and compared overlap. We repeated this simulation 10,000 times. A total of 21 overlapping clones were only observed in 1 (0.01%) of 10,000 simulations, and 22 or more overlapping genes were never observed. These data suggest that the overlap of 41 genes between our BRCA1 tubal epithelium and BRCA1 carcinomas is highly significant and that it did not occur by chance. The concordance of direction of the expression differences in 41 (85%) of 48 overlapping probes also suggests that overlap in differentially expressed genes is nonrandom.

Real-time Quantitative RT-PCR Analysis

Figure 3 shows the Affymetrix expression array results imposed beside the real-time quantitative RT-PCR results for each of the four selected genes (*EFEMP1*, *p57*, *CYP3A5*, and *CSPG5*). For each gene, the real-time quantitative RT-PCR shows a similar expression pattern to the corresponding Affymetrix array.

Clustering Analysis

The 30 samples used to create the *BRCA1* preneoplastic gene signature were subjected to unsupervised hierarchical clustering analysis using all 18,600 probe sets expressed with quality greater than 0.7 (Figure W2). The B1-CA formed a distinct group, but the clustering of the B1-FTocc and WT-FT did not generate any distinct pattern using all expressed probe sets. Interestingly, using only the 41 overlapping probes, the wild-type samples separated into distinct premenopausal and postmenopausal groups, which they did not do when clustering was based on the entire expressed probe set. Of 10 carcinomas evaluated, 6 contained a somatic p53 mutation determined by sequencing p53 exons 4 to 10 (data not shown). Carcinomas 2, 5, 6, 7, 10, and 11 had p53 mutations, whereas carcinomas 1, 3, 9, and 12 were wild-type. The p53 mutation status was not associated with how the carcinomas clustered when considering the 41 overlapping probe sets or all of the 18,600 expressed probes.

Table 2. The 41 Probe Sets Demonstrating Concordant Up-regulation or Down-regulation in Both Comparisons between WT-FT and B1-FTocc or B1-CA.

Affymetrix Probe Set	Gene Name	Gene Symbol	WT-FT *vs* B1-FTocc		WT-FT *vs* B1-CA	
			Fold	*P*	Fold	*P*
Downregulated						
230130_at	Transcribed locus	Unknown	3.7	.0020	3.4	.0054
214078_at	Primary neuroblastoma cDNA	Unknown	3.6	.0003	7.4	.0001
201843_s_at	EGF-containing fibulin-like extracellular matrix protein 1	*EFEMP1*	3.4	.0015	12.4	.0001
205568_at	Aquaporin 9	*AQP9*	2.7	.0021	3.4	.0020
214235_at	Cytochrome P450, family 3, subfamily A	*CYP3A5*	2.7	.0048	5.2	.0023
226228_at	Aquaporin 4	*AQP4*	2.6	.0081	7.7	.0003
213182_x_at	Cyclin-dependent kinase inhibitor 1C (p57, Kip2)	*CDKN1C*	2.5	.0008	3.1	.0014
203710_at	Inositol 1,4,5-triphosphate receptor, type 1	*ITPR1*	2.4	.0004	3.7	.0002
214607_at	p21 (CDKN1A)-activated kinase 3	*PAK3*	2.4	.0070	5.8	.0002
231183_s_at	Jagged 1 (Alagille syndrome)	*JAG1*	2.3	.0002	1.9	.0011
218656_s_at	Lipoma HMGIC fusion partner	*LHFP*	2.3	.0032	2.8	.0083
218717_s_at	Leprecan-like 1	*LEPREL1*	2.3	.0034	2.9	.0034
229480_at	MRNA; cDNA DKFZp686I18116	Unknown	2.1	.0018	2.2	.0002
209506_s_at	Nuclear receptor subfamily 2, group F, member 1	*NR2F1*	2.1	.0036	7.5	.0000
231262_at	Transcribed locus	Unknown	2.1	.0035	3.0	.0088
201497_x_at	Myosin, heavy chain 11, smooth muscle	*MYH11*	2.1	.0068	15.5	.0000
236277_at	Primary neuroblastoma cDNA	Unknown	2.0	.0094	3.5	.0001
201885_s_at	Cytochrome b5 reductase 3	*CYB5R3*	2.0	.0004	1.8	.0066
230233_at	Transcribed locus	Unknown	2.0	.0047	1.9	.0073
1557866_at	Chromosome 9 open reading frame 117	*C9orf117*	2.0	.0013	6.0	.0001
201162_at	Insulin-like growth factor binding protein 7	*IGFBP7*	2.0	.0088	6.1	.0000
201427_s_at	Selenoprotein P, plasma, 1	*SEPP1*	2.0	.0004	2.3	.0010
213451_x_at	Transcribed locus, sim to tenascin XB isoform 1	*TNXB*	1.9	.0072	5.3	.0002
218087_s_at	Sorbin and SH3 domain containing 1	*SORBS1*	1.9	.0082	2.1	.0038
204235_s_at	GULP, engulfment adaptor PTB domain containing 1	*GULP1*	1.9	.0006	4.8	.0000
218718_at	Platelet-derived growth factor C	*PDGFC*	1.9	.0007	1.8	.0059
209575_at	Interleukin 10 receptor, beta	*IL10RB*	1.9	.0010	1.8	.0016
209505_at	Nuclear receptor subfamily 2, group F, member 1	*NR2F1*	1.8	.0096	7.0	.0000
37005_at	Neuroblastoma, suppression of tumorigenicity 1	*NBL1*	1.8	.0066	2.5	.0017
Upregulated						
225857_s_at	Hypothetical LOC388796	*LOC388796*	2.4	.0000	2.6	.0000
238482_at	Kruppel-like factor 7 (ubiquitous)	*KLF7*	2.4	.0020	2.2	.0050
39966_at	Chondroitin sulfate proteoglycan 5	*CSPG5*	2.1	.0030	6.2	.0000
203693_s_at	E2F transcription factor 3	*E2F3*	2.1	.0054	6.4	.0000
1560622_at	CDNA FLJ20196 fis, clone COLF0944	Unknown	2.0	.0038	2.2	.0005
201577_at	Nonmetastatic cells 1, protein (NM23A)	*NME1*	1.9	.0066	3.0	.0002
65588_at	Hypothetical LOC388796	*LOC388796*	1.9	.0000	2.3	.0001
224474_x_at	SMEK homolog 2, suppressor of mek1	*SMEK2*	1.9	.0087	2.0	.0037
212020_s_at	Antigen identified by monoclonal antibody Ki-67	*MKI67*	1.9	.0056	3.5	.0000
224623_at	Transcribed locus, similar THO complex 3	*THOC3*	1.8	.0003	3.2	.0000
1560258_a_at	*Homo sapiens*, clone IMAGE:5590287, mRNA	Unknown	1.8	.0044	3.0	.0000
216262_s_at	TGFB-induced factor homeobox 2	*TGIF2*	1.8	.0094	1.8	.0033

The duplicated samples from the B1-FTocc group were then added to the clustering analysis and subjected to unsupervised hierarchical clustering using the BRCA1 preneoplastic gene signature. As shown in Figure W3, each of the duplicated samples clustered closely with their paired sample even when obtained from the contralateral FT, demonstrating the reproducibility of the expression profile in independent experiments as well as the consistency between paired bilateral fallopian tubes.

To further interrogate the *BRCA1* preneoplastic gene signature, expression profiles from 12 additional B1-FT that had not been used in developing the expression profile were each individually combined with the 30 samples used to create the signature and subjected to unsupervised hierarchical clustering using the *BRCA1* preneoplastic gene signature. Five of the new samples clustered with the B1-FTocc/B1-CA group, whereas seven of the new samples clustered with the WT-FT (Table W2). A representative example of each clustering pattern is shown in Figure W4. This suggested that 5 (42%) of the 12 B1-FT had experienced sufficient molecular disruptions to resemble B1-FTocc or B1-CA samples. Although the remaining seven B1-FT test samples clustered with wild-type fallopian tubes, they always clustered with the group from premenopausal women.

Ki-67 Immunohistochemistry

The Affymetrix array analysis showed a significantly increased expression of *MKI67* (gene for the antigen identified by monoclonal antibody Ki-67) in the B1-FTocc compared with WT-FT ($P = .01$; Figure 4*A*). To confirm the generalizability of the preneoplastic expression pattern in a larger set of wild-type and *BRCA1* histologically normal FT, we evaluated Ki-67 protein expression by immunohistochemistry in a larger set of paraffin-embedded FT specimens. Significantly higher Ki-67 protein expression was identified in fallopian tubes from *BRCA1* mutation carriers than from *BRCA1* wild-type women ($P = .0002$) who did not have cancer (Figure 4*B*). Representative images of low Ki-67 staining in WT-FT (Figure 4*C*) and high Ki-67 staining in B1-FT (Figure 4*D*) are shown.

Discussion

For many years, it was believed that ovarian carcinoma arises from the ovarian surface epithelium or in cortical inclusion cysts in the ovary. In accordance with this belief, most studies assessing disruption of gene expression in ovarian carcinomas have focused on the ovarian surface epithelium and carcinomas within the ovarian tissue [15]. However,

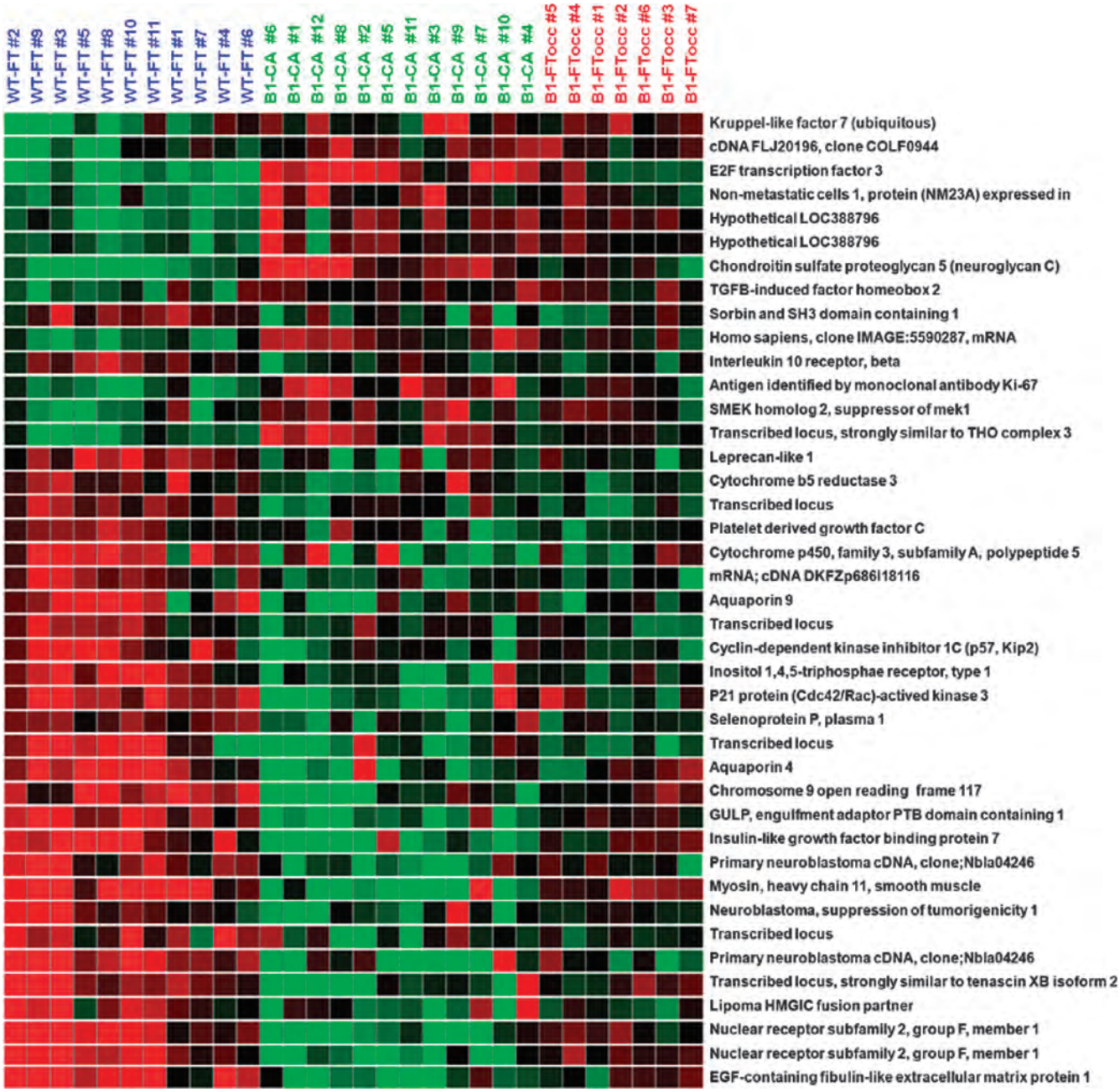


Figure 2. The 41 probe sets in the *BRCA1* preneoplastic profile include several known tumor suppressors and oncogenes.

the relevance of the ovarian surface epithelium has come under increasing scrutiny, and a comprehensive review of the literature regarding the origin of ovarian carcinoma concluded that there is insufficient evidence to support ovarian surface epithelium or inclusion glands as the origin of ovarian carcinomas [16]. In contrast, there has been increasing evidence that many ovarian and primary peritoneal carcinomas arise from neoplastic alterations within the fallopian tube epithelium. This view has been supported by the frequent discovery of occult fallopian tube neoplasms in fallopian tubes removed prophylactically from women at high risk due to hereditary *BRCA1* or *BRCA2* mutations [9–11]. The tubal epithelium in women with *BRCA1* mutations who have up to a 50% lifetime risk of ovarian carcinoma could represent a unique opportunity to study at-risk tissues just before neoplastic transformation. We hypothesized that the epithelium in these high-risk fallopian tubes would express some of the earliest gene disruptions leading to ovarian carcinoma.

Half of all *BRCA1* mutation carriers never develop ovarian carcinoma. This fact could make it difficult to identify a *BRCA1* preneoplastic gene expression profile in normal BRCA1 tubal epithelium in cancer-free *BRCA1* mutation carriers. Our current study is unique because we used histologically normal fallopian tube epithelium from the same fallopian tubes that contained at least a high-grade intraepithelial neoplasm. By using tissues already proven susceptible to neoplastic transformation, we improved our ability to identify a *BRCA1* preneoplastic expression profile in *BRCA1* mutation carriers. We predicted that gene expression differences that precede morphologically identifiable neoplastic transformation should also be present in *BRCA1*-associated ovarian carcinomas. Indeed, 41 of 152 differentially expressed probe sets in

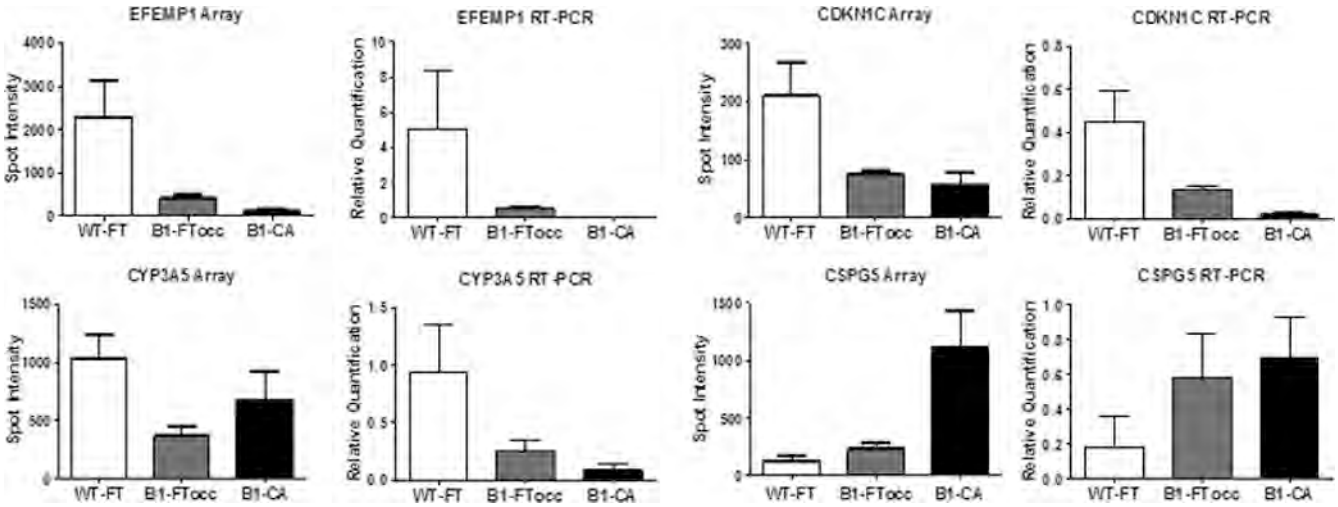


Figure 3. Correlation between expression array and real-time quantitative RT-PCR results: Four genes from the gene signature were selected for validation by RT-PCR with TaqMan assays. Five cases were used from each group (WT-FT, B1-FTocc, and B1-CA). The four genes included *EFEMP1* (EGF-containing fibulin-like extracellular matrix protein 1), *CDKN1C* (cyclin-dependent kinase inhibitor 1C or p57), *CYP3A5* (cytochrome P450, family 3, subfamily A), and *CSPG5* (chondroitin sulfate proteoglycan 5 or neuroglycan C). For each gene, the array expression data are shown beside the corresponding RT-PCR results.

the normal tubal epithelium from *BRCA1* mutation carriers with tubal neoplasia were also similarly differentially expressed in the *BRCA1* carcinomas when compared with tubal epithelium from normal-risk women. Our computer model confirmed that the identified overlap in expression profiles between *BRCA1* tubal epithelium and *BRCA1* carcinoma is highly significant, suggesting that the expression profile that we termed the *BRCA1 preneoplastic signature* represents a true biological phenomenon. The 41 overlapping probe sets represent unique genes altered in

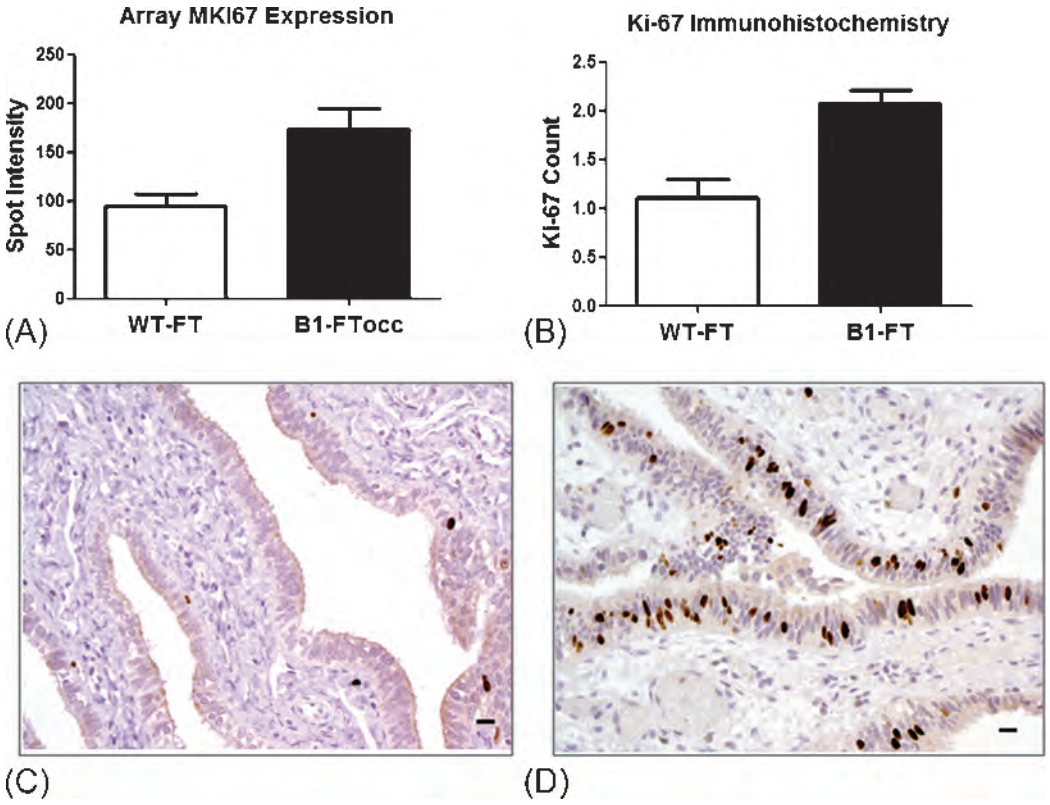


Figure 4. Validation of *MIK67* expression data with Ki-67 immunohistochemistry in 26 WT-FT and 52 B1-FT. (A) *MIK67* gene expression in the 11 WT-FT samples compared with the 7 B1-FTocc samples ($P = .01$). (B) Ki-67 protein expression (brown) in the fallopian tubes from 26 wild-type women compared with fallopian tubes from 52 women with deleterious *BRCA1* mutations ($P = .0002$). Representative images of low Ki-67 expression in a wild-type case (C) and high Ki-67 staining in a *BRCA1*-mutated case (D).

progression from normal fallopian tube epithelium to carcinoma. Furthermore, many of these 41 probe sets represent genes that have been shown to play an important role in cancer biology, such as *EFEMP1*, *CYP3A5*, *CDKN1C*, *NR2F1*, *E2F3*, *MKI67*, *NME1*, and *CSPG5*.

One gene in the *BRCA1* preneoplastic signature overexpressed in BRCA1 FT is the gene encoding the Ki-67 antigen, expressed in the nuclei of proliferating cells. To generalize our findings to other cases from women with known *BRCA1* mutations, we performed immunohistochemistry in a larger series of normal FTs. Consistent with the array data, our pathologists (who were blinded to case designation) identified significantly higher Ki-67 protein expression in FT epithelium of women with *BRCA1* mutations compared to women with negative genetic testing (Figure 4). These data suggest that at least some elements of the *BRCA1* preneoplastic signature are generalizable to BRCA1-mutated FTs without neoplasia. These data suggest that before neoplastic transformation, there exists a higher rate of proliferation in *BRCA1* tubal epithelium, which could increase the opportunity for somatic clonal genetic changes (such as loss of the wild-type allele) and subsequent neoplastic development.

Examples of downregulated probe sets in the *BRCA1* preneoplastic signature include those representing *EFEMP1*, *CDKN1C*, and *NR2F1*. Decreased expression of each of these genes has been implicated in carcinogenesis in a variety of neoplasms. *EFEMP1* (FLBN3) is a member of the fibulin family, a family of secreted glycoproteins with repeated epidermal growth factor domains and a unique C-terminal fibulin-type module. Fibulins mediate cell-to-cell and cell-to-matrix communication within the extracellular matrix [17]. Mutations in *EFEMP1* cause an autosomal-dominant disorder associated with early onset macular degeneration (Doyne honeycomb retinal dystrophy), which has been associated with excessive angiogenesis [18]. *EFEMP1* has antiangiogenic properties and has been shown to inhibit tumor growth in mice. The expression of *EFEMP1* is reduced in many human neoplasms, including ovarian carcinoma [19], and *EFEMP1* is inactivated by promoter methylation in 38% of primary lung carcinomas but not in paired normal lung tissue [20]. The cell cycle regulatory gene *CDKN1C* (p57/Kip2) is an imprinted maternally expressed gene on chromosome 11p15.4. Disruption of *CDKN1C* expression causes the cancer predisposing Beckwith-Wiedemann syndrome [21]. *CDKN1C* has also been implicated as a tumor suppressor gene in a number of human malignant neoplasms including breast, lung, pancreatic, bladder, esophageal, and a variety of hematological and myeloid neoplasms [22,23]. Prostate explants from a *CDKN1C* knockout mouse develop IEN and prostate adenocarcinoma in nude mice, providing the first mouse model that is pathologically identical to human prostate carcinoma [24]. *CDKN1C* dysregulation has not been extensively studied in ovarian carcinoma, but the majority (75%) of sporadic ovarian carcinomas demonstrate reduced CDKN1C protein expression (<10% of tumor cells) using immunohistochemistry [25]. *NR2F1* encodes for the protein chicken ovalbumin upstream promoter transcription factor I (COUP-TF1). COUP-TF1 is a nuclear receptor that has been shown to repress transcription, influence the tumor necrosis factor α signaling pathway [26], and modulate the retinoic acid receptor [27]. In breast carcinoma, decreased expression of COUP-TF1 is associated with the up-regulation of aromatase expression [28]. Decreased expression of COUP-TF1 has also been observed in ovarian and bladder carcinomas [29,30].

Upregulated probe sets in our *BRCA1* preneoplastic signature included *E2F3*, *NME1*, *CSPG5*, and *MKI67*. The *E2F3* gene is a transcription factor that has been implicated in malignant transformation of human lung [31], prostate [32], and bladder carcinomas [33]. Up-regulation of E2F transcription factors has been shown to influence disruptions of the cell cycle in high-grade serous ovarian carcinomas [34], and E2F3 has been used as a biomarker for ovarian carcinoma [35]. The E2F3–Aurora-A axis has been implicated in colorectal cancer [36] and ovarian cancer [37], and recently, *E2F3* has been implicated in the proliferation of ovarian cancer cells through interaction with epidermal growth factor receptor [38]. *NME1* (*NM23*) overexpression has been associated with decreased overall survival in patients with serous ovarian carcinoma [39]. *CSPG5* (neuroglycan C, neuregulin-6) is a growth factor that transactivates the *ErbB2* (HER2/*neu*) oncogene. *CSPG5* is a membrane-anchored chondroitin sulfate proteoglycan that stimulates cell proliferation in a dose-dependent fashion, acts as a specific ligand for ErbB3, and is capable of transactivation of ErbB2 (HER2) [40]. *ErbB2* (HER2/*neu*) is a well-recognized oncogene capable of inducing cellular proliferation and disrupting epithelial cellular polarity. Although *CSPG5* has not been well studied in human malignant neoplasms, *CSPG5* is secreted by neural stem cells, and it promotes its own proliferation in the fetal brain [41].

The traditional clonal model of carcinogenesis states that clonal expansion and neoplastic proliferation stem from genetic disruptions within an individual cell. However, a more contemporary hypothesis called the *epigenetic progenitor model* proposes that before this clonal event, there are global epigenetic alterations in nonneoplastic cell lines that allow the proliferation of cell line–specific stem or progenitor cells. This results in a large population of epigenetically disrupted progenitor cells that could then be affected by an initiating mutation of a key gatekeeper gene in a single cell [42]. An epigenetic progenitor model could explain our ability to identify global alterations of gene expression of key tumor progenitor genes in at-risk epithelium in areas that do not have histologically identifiable neoplastic proliferation. Further epigenetic studies will be necessary to assess this hypothesis in *BRCA1* tubal epithelium.

We assessed whether the tubal expression profile was consistent between various areas of the distal FT by performing unsupervised hierarchical clustering using independent samples from three of the B1-FTocc cases (two from the contralateral tube). Regardless of whether the duplicated samples were created from the ipsilateral or contralateral fallopian tube, all three duplicates clustered immediately adjacent to their corresponding sample when considering the preneoplastic gene signature (Figure W3). This suggests that the gene disruptions we observed in high-risk fallopian tubes represent a global field effect that affects bilateral fallopian tubes in patients with *BRCA1* mutations.

p53-immunopositive foci have been frequently observed in tubal epithelium of both high-risk and normal-risk women [3,13]. We made no effort to select p53-positive cells to derive the BRCA1 preneoplastic expression profile. The resulting expression profile did not seem to be driven by p53 because the expression profiles from the p53 wild-type carcinomas were not distinct from the p53 mutant carcinomas when just considering these genes. The probe sets on the Affymetrix array representing p53 showed minimal signal regardless of BRCA1 status. This is not surprising because p53 foci are generally small, occurring in as few as 10 cells and, consequently, would only be present in a small fraction of the cells that we used for expression array analysis. p53 foci likely represent a clonal event (somatic mutation) in a small subset of tubal epithelial cells. The fact that we can detect differences in expression profiles of *BRCA1* tubal epithelium despite not selecting for p53 foci implies that global alterations in gene expression including *MKI67* (Ki-67 protein) occur even in cells that do not have p53 alterations or mutation. These data support an alternate model in which

global alterations (including increased Ki-67) precede somatic clonal events such as p53 mutation in p53 foci [13].

Three of the seven B1-FTocc samples available from our tissue bank were collected from premenopausal women. To equalize the menopausal status in our three groups, we specifically included cases in the WT-FT and B1-CA groups that were premenopausal at the time of surgery. When unsupervised hierarchical clustering was performed using the preneoplastic gene signature, the four premenopausal WT-FT cases formed a distinct group from the seven postmenopausal WT-FT cases. Interestingly, when the 12 additional FT-B1 samples (which were all postmenopausal) were subjected to clustering analysis, the 7 samples that clustered with the WT-FT group always clustered with the premenopausal WT-FT. It seems that BRCA1-mutated fallopian tubes maintain a gene expression profile that is more similar to premenopausal tissue, even without the stimulation of the premenopausal hormonal milieu. Our group has recently demonstrated that proliferation in WT-FT as measured by Ki-67 protein expression decreases with age, but Ki-67 expression is maintained at a higher level with a less marked decrease with age in women with *BRCA1* mutation [13]. Therefore, both protein expression and expression profiling suggest that *BRCA1* fallopian tube epithelium maintains a premenopausal proliferative phenotype. Overall, 5 (42%) of the 12 additional B1-FT samples clustered with the B1-FTocc/B1-CA group based on the *BRCA1* preneoplastic signature. This closely reflects the percentage of women with *BRCA1* mutations who will go on to develop ovarian carcinoma [8]. Interestingly, the samples that clustered with the B1-FTocc/B1-CA group had higher Ki-67 staining.

There has only been one published study by Tone et al. [43] looking at differential gene expression profiles from *BRCA1*-mutated fallopian tube epithelium and fallopian tube/ovarian carcinomas, which was designed differently from our study. These investigators analyzed fallopian tube epithelium only from premenopausal women, included both *BRCA1* and *BRCA2* mutation carriers, and focused on fallopian tubes without associated carcinoma, as opposed to our strategy of microdissecting epithelium from fallopian tubes containing at least high-grade intraepithelial neoplasm. Furthermore, their carcinoma group included sporadic fallopian tube and ovarian carcinomas, whereas we compared expression profiles specifically to BRCA1-mutated carcinomas. We felt it was important to separate BRCA1 from BRCA2 fallopian tube epithelium given that ovarian carcinomas have distinct different expression profiles according to whether they have a *BRCA1* or *BRCA2* mutation [44]. They observed that *BRCA1*-mutated fallopian tubes collected during the luteal phase of the menstrual cycle were more likely to cluster with the carcinoma samples. They hypothesized that the hormonal environment of the luteal phase causes distinct changes in high-risk fallopian tubes resulting in similar gene expression to carcinoma tissue. Because of the different approaches between this study and our current study, it is difficult to compare the specific genes identified. However, both studies suggest that fallopian tube epithelium from *BRCA1* mutation carriers is susceptible to disruption in gene expression, which causes histologically normal fallopian tube tissue to exhibit gene expression resembling carcinoma.

By analyzing gene expression from histologically normal fallopian tube epithelium isolated from *BRCA1*-mutated fallopian tubes containing early neoplasms, we have identified a potential *BRCA1* preneoplastic gene expression signature for *BRCA1* serous carcinoma. This gene signature may include some of the earliest disruptions in gene expression leading to the development of serous ovarian carcinoma. Further validation will be necessary to determine which of the genes from this signature are critical in this process and to identify the mechanisms of gene expression alterations. The fact that these genes are disrupted in the fallopian tube tissue before the development of invasive carcinoma could make them useful targets for chemoprevention or early detection of ovarian carcinoma.

Acknowledgments

The authors thank Mary-Claire King for her support and guidance in this project, the staff at the Centre for Array Technology for their assistance with processing gene expression arrays, Danbin Xu for guidance with real-time PCR analysis, Robert Vessella and Colm Morrissey for assistance with LCM, Lawrence True for guidance with tissue preparation, and Peter Nelson and Ilsa Coleman for LCM and expression array protocols.

References

[1] Cai KQ, Klein-Szanto A, Karthik D, Edelson M, Daly MB, Ozols RF, Lynch HT, Godwin AK, and Xu XX (2006). Age-dependent morphological alterations of human ovaries from populations with and without *BRCA* mutations. *Gynecol Oncol* **103**, 719–728.

[2] Schlosshauer PW, Cohen CJ, Penault-Llorca F, Miranda CR, Bignon YJ, Dauplat J, and Deligdisch L (2003). Prophylactic oophorectomy: a morphologic and immunohistochemical study. *Cancer* **98**, 2599–2606.

[3] Crum CP, Drapkin R, Miron A, Ince TA, Muto M, Kindelberger DW, and Lee Y (2007). The distal fallopian tube: a new model for pelvic serous carcinogenesis. *Curr Opin Obstet Gynecol* **19**, 3–9.

[4] Kmet LM, Cook LS, and Magliocco A (2003). A review of p53 expression and mutation in human benign, low malignant potential, and invasive epithelial ovarian tumors. *Cancer* **97**, 389–404.

[5] Stern J, Buscema J, Parmley T, Woodruff JD, and Rosenshein NB (1981). Atypical epithelial proliferations in the fallopian tube. *Am J Obstet Gynecol* **140**, 309–312.

[6] Yanai-Inbar I and Silverberg SG (2000). Mucosal epithelial proliferation of the fallopian tube: prevalence, clinical associations, and optimal strategy for histopathologic assessment. *Int J Gynecol Pathol* **19**, 139–144.

[7] Kindelberger DW, Lee Y, Miron A, Hirsch MS, Feltmate C, Medeiros F, Callahan MJ, Garner EO, Gordon RW, Birch C, et al. (2007). Intraepithelial carcinoma of the fimbria and pelvic serous carcinoma: evidence for a causal relationship. *Am J Surg Pathol* **31**, 161–169.

[8] King MC, Marks JH, and Mandell JB (2003). Breast and ovarian cancer risk due to inherited mutations in *BRCA1* and *BRCA2*. *Science* **302**, 643–646.

[9] Finch A, Shaw P, Rosen B, Murphy J, Narod SA, and Colgan TJ (2006). Clinical and pathologic findings of prophylactic salpingo-oophorectomies in 159 *BRCA1* and *BRCA2* carriers. *Gynecol Oncol* **100**, 58–64.

[10] Lamb JD, Garcia RL, Goff BA, Paley PJ, and Swisher EM (2006). Predictors of occult neoplasia in women undergoing risk-reducing salpingo-oophorectomy. *Am J Obstet Gynecol* **194**, 1702–1709.

[11] Leeper K, Garcia R, Swisher E, Goff B, Greer B, and Paley P (2002). Pathologic findings in prophylactic oophorectomy specimens in high-risk women. *Gynecol Oncol* **87**, 52–56.

[12] Lee Y, Miron A, Drapkin R, Nucci MR, Medeiros F, Saleemuddin A, Garber J, Birch C, Mou H, Gordon RW, et al. (2007). A candidate precursor to serous carcinoma that originates in the distal fallopian tube. *J Pathol* **211**, 26–35.

[13] Norquist BM, Garcia RL, Allison KH, Jokinen CH, Kernochan LE, Pizzi CC, Barrow BJ, Goff BA, and Swisher EM (in press). The molecular pathogenesis of hereditary ovarian carcinoma: alterations in the tubal epithelium of women with *BRCA1* and *BRCA2* mutations. *Cancer*.

[14] Galic V, Willner J, Wollan M, Garg R, Garcia R, Goff BA, Gray HJ, and Swisher EM (2007). Common polymorphisms in TP53 and MDM2 and the relationship to TP53 mutations and clinical outcomes in women with ovarian and peritoneal carcinomas. *Genes Chromosomes Cancer* **46**, 239–247.

[15] Fehrmann RS, Li XY, van der Zee AG, de Jong S, Te Meerman GJ, de Vries EG, and Crijns AP (2007). Profiling studies in ovarian cancer: a review. *Oncologist* **12**, 960–966.

[16] Bell DA (2005). Origins and molecular pathology of ovarian cancer. *Mod Pathol* **18**(suppl 2), S19–S32.

[17] Gallagher WM, Currid CA, and Whelan LC (2005). Fibulins and cancer: friend or foe? *Trends Mol Med* **11**, 336–340.

[18] Stone EM, Lotery AJ, Munier FL, Héon E, Piguet B, Guymer RH, Vandenburgh K, Cousin P, Nishimura D, Swiderski RE, et al. (1999). A single *EFEMP1* mutation associated with both Malattia Leventinese and Doyne honeycomb retinal dystrophy. *Nat Genet* **22**, 199–202.

[19] Albig AR, Neil JR, and Schiemann WP (2006). Fibulins 3 and 5 antagonize tumor angiogenesis *in vivo*. *Cancer Res* **66**, 2621–2629.

[20] Yue W, Dacic S, Sun Q, Landreneau R, Guo M, Zhou W, Siegfried JM, Yu J, and Zhang L (2007). Frequent inactivation of *RAMP2*, *EFEMP1* and *Dutt1* in lung cancer by promoter hypermethylation. *Clin Cancer Res* **13**, 4336–4344.

[21] Hatada I, Ohashi H, Fukushima Y, Kaneko Y, Inoue M, Komoto Y, Okada A, Ohishi S, Nabetani A, Morisaki H, et al. (1996). An imprinted gene $p57^{KIP2}$ is mutated in Beckwith-Wiedemann syndrome. *Nat Genet* **14**, 171–173.

[22] Larson PS, Schlechter BL, King CL, Yang Q, Glass CN, Mack C, Pistey R, de Las Morenas A, and Rosenberg CL (2008). *CDKN1C/*$p57^{kip2}$ is a candidate tumor suppressor gene in human breast cancer. *BMC Cancer* **8**, 68.

[23] Kikuchi T, Toyota M, Itoh F, Suzuki H, Obata T, Yamamoto H, Kakiuchi H, Kusano M, Issa JP, Tokino T, et al. (2002). Inactivation of $p57^{KIP2}$ by regional promoter hypermethylation and histone deacetylation in human tumors. *Oncogene* **21**, 2741–2749.

[24] Jin RJ, Lho Y, Wang Y, Ao M, Revelo MP, Hayward SW, Wills ML, Logan SK, Zhang P, and Matusik RJ (2008). Down-regulation of $p57^{Kip2}$ induces prostate cancer in the mouse. *Cancer Res* **68**, 3601–3608.

[25] Khouja MH, Baekelandt M, Nesland JM, and Holm R (2007). The clinical importance of Ki-67, p16, p14, and p57 expression in patients with advanced ovarian carcinoma. *Int J Gynecol Pathol* **26**, 418–425.

[26] Zhang LJ, Liu X, Gafken PR, Kioussi C, and Leid M (2008). A chicken ovalbumin upstream transcription factor 1 (COUP-TF1) complex represses expression of the gene encoding tumor necrosis factor α–induced protein 8 (*TNFAIP8*). *J Biol Chem* **284**, 6156–6168.

[27] Lin B, Chen GQ, Xiao D, Kolluri SK, Cao X, Su H, and Zhang XK (2000). Orphan receptor COUP-TF is required for induction of retinoic acid receptor beta, growth inhibition, and apoptosis by retinoic acid in cancer cells. *Mol Cell Biol* **20**, 957–970.

[28] Chen S, Jingjing Y, Ikuko K, Yoshiyuki K, and Zhou D (2005). Positive and negative transcriptional regulation of aromatase expression in human breast cancer tissue. *J Steroid Biochem Mole Biol* **95**, 17–23.

[29] Ham WS, Lee JH, Yu HS, and Choi YD (2008). Expression of chicken ovalbumin upstream promoter-transcription factor 1 (COUP-TF1) in bladder transitional cell carcinoma. *Urology* **72**, 921–926.

[30] Damiao RDS, Oshima CTF, Stavale JN, and Goncalves WJ (2007). Analysis of the expression of estrogen receptor, progesterone receptor and chicken ovalbumin upstream promoter-transcription factor 1 in ovarian epithelial cancers and normal ovaries. *Oncol Rep* **18**, 25–32.

[31] Cooper CS, Nicholson AG, Foster C, Dodson A, Edwards S, Fletcher A, Roe T, Clark J, Joshi A, Norman A, et al. (2006). Nuclear overexpression of the E2F3 transcription factor in human lung cancer. *Lung Cancer* **54**, 155–162.

[32] Foster CS, Falconer A, Dodson AR, Norman AR, Dennis N, Fletcher A, Southgate C, Dowe A, Dearnaley D, Jhavar S, et al. (2004). Transcription factor E2F3 overexpressed in prostate cancer independently predicts clinical outcome. *Oncogene* **23**, 5871–5879.

[33] Feber A, Clark J, Goodwin G, Dodson AR, Smith PH, Fletcher A, Edwards S, Flohr P, Falconer A, Roe T, et al. (2004). Amplification and overexpression of E2F3 in human bladder cancer. *Oncogene* **23**, 1627–1630.

[34] Meyer TD, Bijsmans IT, Van de Vijver KK, Bekaert S, Oosting J, Van Criekinge W, van Engeland M, and Sieben NL (2009). E2Fs mediate a fundamental cell-cycle deregulation in high-grade serous ovarian carcinomas. *J Pathol* **217**, 14–20.

[35] Lu KH, Patterson AP, Wang L, Marquez RT, Atkinson EN, Baggerly KA, Ramoth LR, Rosen DG, Liu J, Hellstrom I, et al. (2004). Selection of potential markers for epithelial ovarian cancer with gene expression arrays and recursive descent partition analysis. *Clin Cancer Res* **10**, 3291–3300.

[36] Baba Y, Nosho K, Shima K, Irahara N, Kure S, Toyoda S, Kirkner GJ, Goel A, Fuchs CS, and Ogino A (2009). Aurora-A expression is independently associated with chromosomal instability in colorectal cancer. *Neoplasia* **11**, 418–425.

[37] He L, Yang H, Ma Y, Pledger WJ, Cress WD, and Cheng JQ (2008). Identification of Aurora-A as a direct target of E2F3 during G_2/M cell cycle progression. *J Biol Chem* **283**, 31012–31020.

[38] Reimer D, Hubalek M, Riedle S, Skvortsov S, Erdel M, Conci N, Fiegl H, Muller-Holzner E, Marth C, Illmensee K, et al. (2010). E2F3a is critically involved in epidermal growth factor receptor–directed proliferation in ovarian cancer. *Cancer Res* **70**, 4613–4623.

[39] Youn BS, Kim DS, Kim JW, Kim YT, Kang S, and Cho NH (2008). NM23 as a prognostic biomarker in ovarian serous carcinoma. *Mod Pathol* **21**, 885–892.

[40] Kinugasa Y, Ishiguro H, Tokita Y, Oohira A, Ohmoto H, and Higashiyama S (2004). Neuroglycan C, a novel member of the neuregulin family. *Biochem Biophys Res Commun* **321**, 1045–1049.

[41] Ida M, Shuo T, Hirano K, Tokita Y, Nakanishi K, Matsui F, Aono S, Fujita H, Fujiwara Y, Kaji T, et al. (2006). Identification and functions of chondroitin sulfate in the milieu of neural stem cells. *J Biol Chem* **281**, 5982–5991.

[42] Feinberg AP, Ohlsson R, and Henikoff S (2006). The epigenetic progenitor origin of human cancer. *Nat Rev Genet* **7**, 21–33.

[43] Tone AA, Begley H, Sharma M, Murphy J, Rosen B, Brown TJ, and Shaw PA (2008). Gene expression profiles of luteal phase fallopian tube epithelium from BRCA mutation carriers resemble high-grade serous carcinomas. *Clin Cancer Res* **14**, 4067–4078.

[44] Jazaeri AA, Yee CJ, Sotiriou C, Brantley KR, Boyd J, and Liu ET (2002). Gene expression profiles of BRCA1-linked, BRCA2-linked, and sporadic ovarian cancers. *J Natl Cancer Inst* **94**, 990–1000.

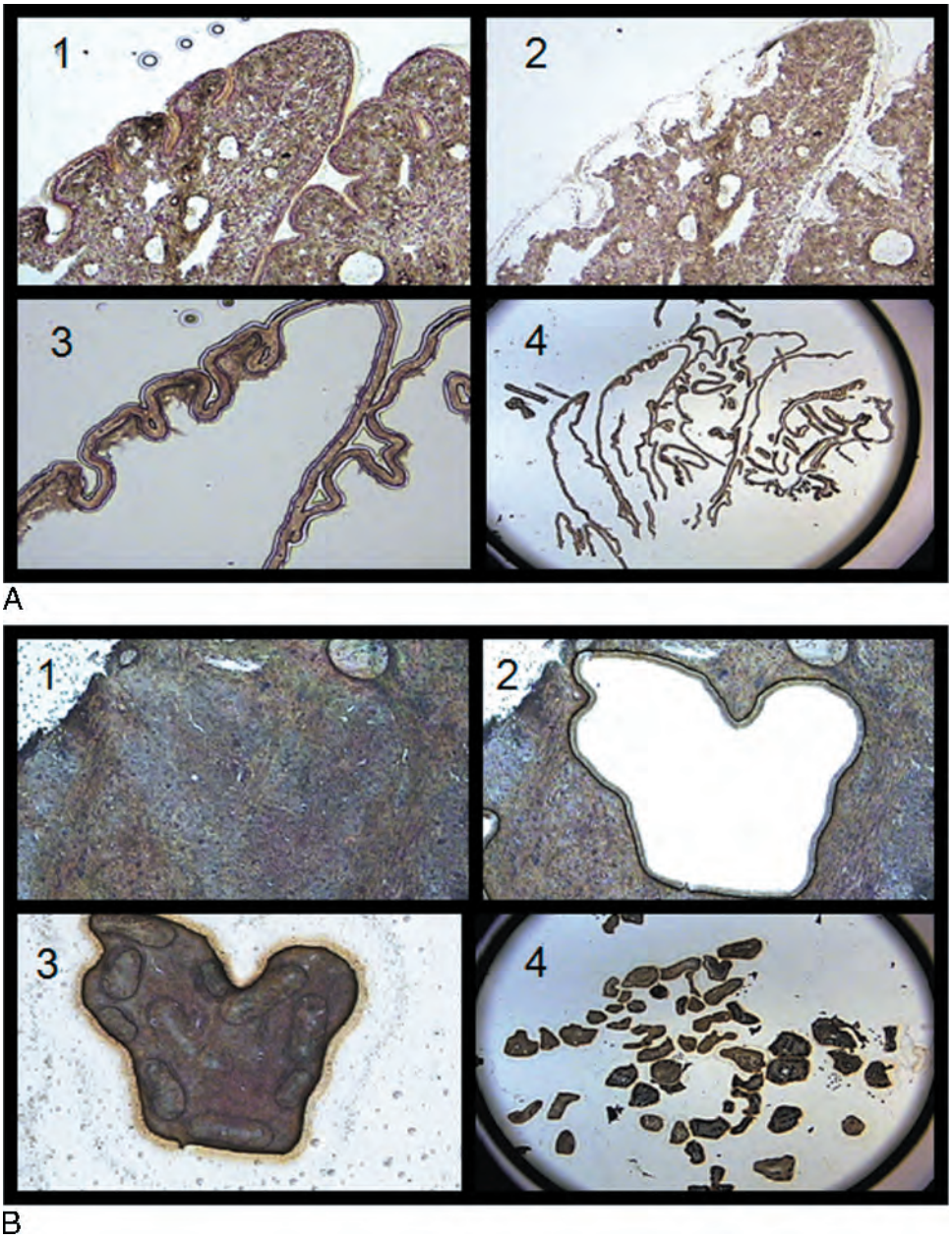


Figure W1. Illustration of LCM. (A) Fallopian tube tissue: (1) fixed slide immediately before performing LCM, (2) slide after LCM and removal of tissue, (3) fallopian tube epithelial tissue adherent to collection cap, and (4) collection cap coated with fallopian tube epithelial tissue at completion. (B) Tumor tissue: same steps (1-4) are shown.

Table W1. Duplicated Samples Used to Interrogate the Gene Expression Signature.

Unique Identifier	Corresponding Case	Tissue Block Used for LCM	BRCA1 Status	Menopausal Status
B1-FTocc no. 1 – DUP	B1-FTocc no. 1	Ipsilateral FT	B1.3109insAA	Pre
B1-FTocc no. 2 – DUP	B1-FTocc no. 2	Contralateral FT	(120A>G)	Post
B1-FTocc no. 6 – DUP	B1-FTocc no. 6	Contralateral FT	B1.C61G	Post

For each of these patients, an independent section of fallopian tube was subjected to RNA isolation, amplification, and array creation.

Table W2. Twelve Additional Fallopian Tube Samples Used to Interrogate the *BRCA1* Preneoplastic Gene Signature.

Case Identifier	Age (years)	Menopausal Status	*BRCA1* Status	Clustering Group
B1-FT no. 1	39	Post	B1.IVS5-11 T>G	WT-FT
B1-FT no. 2	43	Post	B1.C61G	WT-FT
B1-FT no. 3	45	Post	B1.5677insA	B1-FTocc
B1-FT no. 4	46	Post	B1.13+1 G to A	WT-FT
B1-FT no. 5	48	Post	B1.975delAG	WT-FT
B1-FT no. 6	49	Post	B1.3124delA	B1-FTocc
B1-FT no. 7	50	Post	B1.3878insT	WT-FT
B1-FT no. 8	51	Post	B1.Q1200X	B1-FTocc
B1-FT no. 9	52	Post	B1.120A>G(M1V)	B1-FTocc
B1-FT no. 10	53	Post	B1.5385insC	WT-FT
B1-FT no. 11	59	Post	B1.del exon 17	B1-FTocc
B1-FT no. 12	62	Post	B1.R1699W	WT-FT

Each fallopian tube was collected at the time of prophylactic salpingo-oophorectomy from a patient with a known deleterious BRCA1 mutations. None of these samples were used in the derivation of the *BRCA1* preneoplastic gene signature.

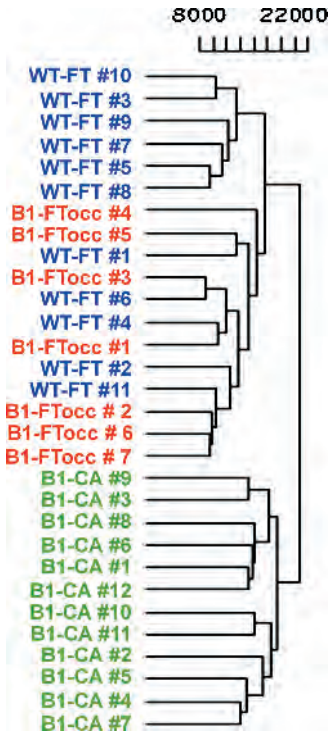


Figure W2. Unsupervised hierarchical clustering of cases using all 18,600 probe sets expressed with adequate quality on the arrays. B1-CA formed a distinct group from the fallopian tube samples.

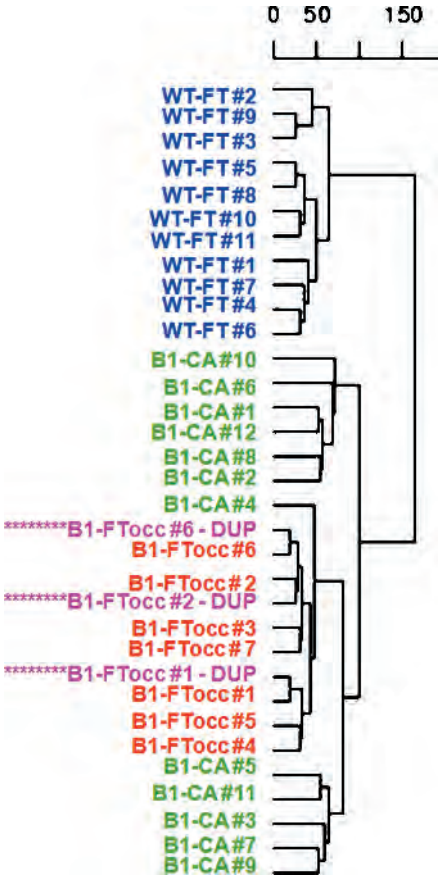


Figure W3. Using the same cases analyzed to create the gene signature, duplicate sections were made, and the protocol was repeated to assess validity. For case no. 1, the ipsilateral FT was used, and for case nos. 2 and 6, the contralateral FT was used. Frozen tissue was subjected *independently* to sectioning, LCM, and RNA amplification. These duplicated samples were then subjected to unsupervised hierarchical clustering using the original 41 probe set gene signature. Clustering of the three duplicated samples (B1-FTocc DUP) shows that duplicates cluster near regardless of whether they are taken from the ipsilateral or contralateral fallopian tube.

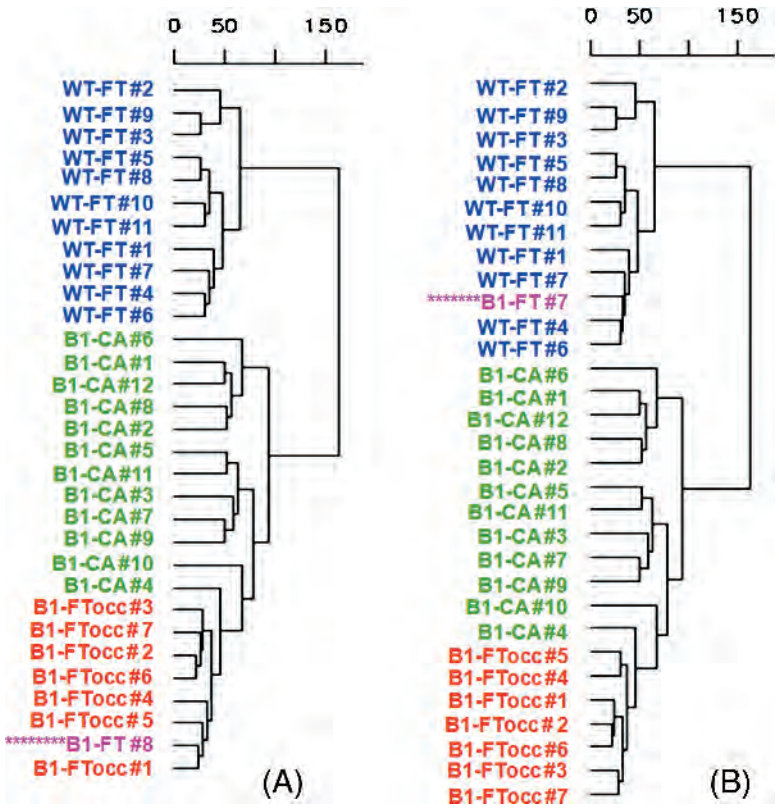


Figure W4. Representative examples of clustering with the independent B1-FT test samples using the 41 probe set gene signature. (A) B1-FT no. 8 clustered with the B1-FTocc/B1-CA group, whereas (B) B1-FT no. 7 clustered with the WT-FT group.

Epigenetics & Chromatin

RESEARCH Open Access

Allele-specific transcriptional elongation regulates monoallelic expression of the *IGF2BP1* gene

Brandon J Thomas[1], Eric D Rubio[1], Niklas Krumm[2], Pilib Ó Broin[3,4], Karol Bomsztyk[5], Piri Welcsh[6], John M Greally[7,8], Aaron A Golden[7] and Anton Krumm[1,9*]

Abstract

Background: Random monoallelic expression contributes to phenotypic variation of cells and organisms. However, the epigenetic mechanisms by which individual alleles are randomly selected for expression are not known. Taking cues from chromatin signatures at imprinted gene loci such as the insulin-like growth factor 2 gene 2 (*IGF2*), we evaluated the contribution of CTCF, a zinc finger protein required for parent-of-origin-specific expression of the *IGF2* gene, as well as a role for allele-specific association with DNA methylation, histone modification and RNA polymerase II.

Results: Using array-based chromatin immunoprecipitation, we identified 293 genomic loci that are associated with both CTCF and histone H3 trimethylated at lysine 9 (H3K9me3). A comparison of their genomic positions with those of previously published monoallelically expressed genes revealed no significant overlap between allele-specifically expressed genes and colocalized CTCF/H3K9me3. To analyze the contributions of CTCF and H3K9me3 to gene regulation in more detail, we focused on the monoallelically expressed *IGF2BP1* gene. *In vitro* binding assays using the CTCF target motif at the *IGF2BP1* gene, as well as allele-specific analysis of cytosine methylation and CTCF binding, revealed that CTCF does not regulate mono- or biallelic *IGF2BP1* expression. Surprisingly, we found that RNA polymerase II is detected on both the maternal and paternal alleles in B lymphoblasts that express *IGF2BP1* primarily from one allele. Thus, allele-specific control of RNA polymerase II elongation regulates the allelic bias of *IGF2BP1* gene expression.

Conclusions: Colocalization of CTCF and H3K9me3 does not represent a reliable chromatin signature indicative of monoallelic expression. Moreover, association of individual alleles with both active (H3K4me3) and silent (H3K27me3) chromatin modifications (allelic bivalent chromatin) or with RNA polymerase II also fails to identify monoallelically expressed gene loci. The selection of individual alleles for expression occurs in part during transcription elongation.

Background

Allele-specific gene expression is an integral component of cellular programming and development and contributes to the diversity of cellular phenotypes [1,2]. Allelic differences in gene expression are mediated by either parent-of-origin-specific selection (imprinting) or stochastic selection of alleles for activation and/or silencing. The importance of genomic imprinting has recently been highlighted by RNA sequencing studies that demonstrated widespread allelic differences in gene expression in mouse brain affecting more than 1,300 genes [3]. The extent of sex- and stage-specific expression of individual alleles emphasizes the essential role of allelic transcriptional regulation in development. In addition to the extensive occurrence of imprinted parent-of-origin-specific expression, gene expression patterns of clonal cell populations are also modified by random or stochastic silencing of either the maternal or paternal allele. Well-known loci displaying allele-specific expression include odorant receptor genes, immunoglobulins and various receptor proteins [4-6]. Additionally, previous large-scale studies have provided new data demonstrating that parent-of-origin-specific expression is employed much more frequently than previously

* Correspondence: akrumm@u.washington.edu
[1]Institute for Stem Cell and Regenerative Medicine, University of Washington School of Medicine, 815 Mercer St., Seattle, WA 98109, USA
Full list of author information is available at the end of the article

thought [7]. These new findings illustrate the scale and complexity of genomic allele-specific expression. However, the precise molecular mechanism underlying the allelic bias in gene expression is not very well understood.

The best-characterized locus with strict monoallelic imprinted gene expression is the region containing the insulin-like growth factor 2 (*IGF2*) and *H19* genes [8]. The regulation of this locus relies on the imprinting control region (ICR), which acquires DNA methylation on the paternal allele during normal development of the male germline. Methylation of cytosines at the ICR inhibits binding of the zinc finger protein CTCF to the paternal allele, preventing its role as an insulator and allowing long-range interactions of the *IGF2* promoter with enhancer elements downstream of the *H19* gene [9-11]. In contrast, the unmethylated ICR on the maternal allele recruits CTCF, effectively preventing promoter-enhancer interactions and maintaining repression of the maternal *IGF2* gene.

The well-documented requirement of CTCF for imprinted expression at the *IGF2*/*H19* gene locus is thought to result from its role in establishing and/or maintaining long-distance interactions between regulatory elements [12]. Allele-specific binding of CTCF to the ICR has long been known to be essential for the formation of chromatin loops. While the precise mechanism of CTCF's role in long-distance chromatin interactions remains unknown, several studies have provided a rationale for the differential expression of the maternal and paternal *IGF2* gene by revealing an interaction of CTCF with cohesin, a protein complex known for its requirement during sister chromatid cohesion in mitosis [13-16]. Chromosome conformation capture experiments in combination with RNA interference assays recently confirmed the CTCF and cohesin-dependent formation of higher-order chromatin structures at the *IGF2*/*H19* and other gene loci [17-19].

In addition to DNA methylation, histone modifications also contribute to the maintenance of allele-specific expression. DNA methylation of ICRs is accompanied by repressive histone markers, including histone H3 trimethylated at lysine 9 (H3K9me3). In contrast, the unmethylated allele is characterized by permissive histone markers, including histone H3 trimethylated at lysine 4 [20]. Colocalization of epigenetic markers including DNA methylation and histone H3 dimethylated at lysine 9 has been exploited to identify epigenetically distinct parental alleles. Chromosomal regions displaying overlaps of euchromatin and heterochromatin-specific markers have been enriched for known imprinted genes [21].

Despite the importance of monoallelic expression in cellular development and differentiation, little is known about the establishment and maintenance of random monoallelic expression. The link between allele-specific binding of CTCF and monoallelic expression of the *IGF2* gene prompted us to test whether the presence of CTCF and H3K9me3 specifies a chromatin arrangement which demarcates random monoallelically expressed alleles. Using array-based chromatin immunoprecipitation (ChIP-chip), we identified 293 loci displaying these chromatin markers. We selected the *IGF2BP1* gene locus to further examine whether the presence of CTCF and H3K9me3 comprises a necessary chromatin arrangement for a specific expression profile analogous to the monoallelic behavior observed at the *IGF2*/*H19* locus. Surprisingly, colocalization of CTCF and H3K9me3 does not provide a reliable measure of monoallelic binding of CTCF at the *IGF2BP1* gene. Our studies included allele-specific sequencing of immunoprecipitated chromatin to demonstrate that chromatin at each *IGF2BP1* allele is bivalent. Importantly, both alleles recruit RNA polymerase II, suggesting that silencing of one *IGF2BP1* allele occurs after transcription initiation. By establishing which epigenetic configurations are involved in governing monoallelic gene expression, we will broaden the understanding of epigenetic mechanisms as they relate to cancer progression and cellular differentiation.

Results

Colocalization of CTCF and H3K9me3 in the human genome

Allele-specific binding of CTCF to the ICR regulates parent-of-origin-specific expression of the *IGF2* gene and correlates with differential cytosine methylation and the presence of H3K9me3 [9-11]. We carried out a large-scale survey to identify genomic sites with chromatin markers similar to those at the ICR of the *IGF2*/*H19* locus. Using ChIP-chip, we identified CTCF binding sites by tiling through the nonrepetitive portion of the genome in 100-bp intervals. Genomic sites bound by CTCF were assembled on a condensed array set that tiled through 9,823 sites using overlapping probes, and replicate ChIP experiments were performed. By using conservative criteria (positive signal in three replicates; $P < 0.05$) in this analysis, we identified 8,462 loci that interact with CTCF. To identify the subset of sites that associate with both CTCF and H3K9me3, we tested the association of these 8,462 loci with H3K9me3 using the condensed DNA array set. These analyses revealed 293 loci that are both bound by CTCF and marked by H3K9me3 (Table S1 in Additional file 1) (distances of CTCF and H3K9me3 peaks < 500 bp). Of the 293 loci, 115 directly mapped to coding regions. Of the remaining loci (174 of 293), the majority (147 loci) were located in intergenic regions at a distance > 10 kb to

the nearest 5' end of known genes. Only 27 loci mapped to promoter regions. Overall, 40% of the CTCF/H3K9me3 loci mapped to intergenic regions, 51% mapped to intragenic domains and 9% mapped to promoter regions, a distribution similar to that of the 8,462 CTCF loci (44%, 51% and 10% respectively). Notably, the CTCF-regulated *IGF2*/*H19* locus is included in the subset of 293 loci (Figure S1 in Additional file 2), suggesting that our experimental approach may be useful for the identification of similarly expressed genes.

IGF2BP1 alleles are stochastically expressed in human B cells

Genes classified as "monoallelically expressed" encompass both imprinted genes, such as the *IGF2* gene, where monoallelic expression is regulated in a parent-of-origin-specific manner, and stochastic loci, where individual alleles are randomly selected for expression independent of parental origin. In recent studies in which allele-specific transcription was assessed in several human cell lines, more than 300 (7.5%) of 4,000 human genes examined were subject to random monoallelic expression, with a majority of the latter being capable of biallelic expression [7].

To examine whether CTCF binding at sites marked by H3K9me3 is indicative of monoallelic expression, we first compared the genomic positions of our 293 loci with the list of genes expressed in a random allele-specific manner. Only a small number of genes (8 of 293 loci) were common to both the monoallelically expressed cohort described by Gimelbrant *et al*. [7] and our CTCF/H3K9me3 set of ChIP-chip binding loci.

To further examine the correlation between CTCF/H3K9me3 and monoallelic expression, we selected 12 genes located near one of the 293 CTCF/H3K9me3 sites (*DIAPH1*, *FUS1*, *PKP1*, *ARFGAP2*, *PCDHGA*, *MTHFR*, *LAIR1*, *GPR3*, *ARMET*, *NPR1*, *NHLRC1* and *IGF2BP1*) to search lymphoblastoid cell lines (LCLs) derived from a pedigree from the Center d'Etude du Polymorphisme Humaine (CEPH) for SNPs in exonic and 3'-UTR regions. The monoclonality of LCLs was confirmed by analysis of their immunoglobulin heavy chain (IgH) gene rearrangement (Figure S2 in Additional file 2) [22]. Sequencing of genomic DNA (gDNA) and cDNA of LCLs identified the insulin-like growth factor binding protein gene *IGF2BP1* as the only candidate gene expressed from only one allele (Table S2 in Additional file 1). *IGF2BP1* is an RNA-binding protein that regulates transcript stability and translation of the imprinted *IGF2* gene [23]. In addition, *IGF2BP1* binds to H19, MYC and β-TrCP1 mRNA to regulate message half-life, localization and translation of RNA, suggesting that the regulation of *IGF2BP1* expression may affect disease and development [24,25]. We focused on *IGF2BP1* to examine the contribution of CTCF and H3K9me3 markers colocalized at intron 5 to allele-specific expression (Figure 1).

Sequencing of gDNA identified 10 individuals that were heterozygous at SNP rs11655950 in the 3'-UTR of *IGF2BP1* (Figure 2A). All heterozygous SNPs were subsequently typed in cDNA. A comparison of the transcriptome-derived genotypes to genomic genotypes indicated that six individuals expressed *IGF2BP1* primarily from only one allele. In contrast, four individuals were found to express both *IGF2BP1* alleles (Figure 2A). SNP determination for genomic and cDNA for CEPH family 1331 was confirmed by allelic discrimination assays based on fluorogenic probes (TaqMan allelic discrimination assay; Applied Biosystems, Foster City, CA, USA), which yielded identical results (Figure S3 in Additional file 2). The TaqMan allelic discrimination assay, a real-time PCR based approach, yields a scatterplot of genotypes capable of quantitatively detecting a range of 1:1 and 1:5 ratios of individual alleles in DNA mixtures at SNP rs11655950 (Figure S4 in Additional file 2). Individuals GM7033 and GM6989 were found to express the paternally inherited *IGF2BP1* allele, while GM7030 and GM7005 were found to express the maternally inherited allele (Figure 2). Individuals GM7007 and GM7016 also exhibited monoallelic expression of *IGF2BP1*, but we were unable to identify the mode of expression because of the limited pedigree. These data indicate that monoallelic expression at the *IGF2BP1* gene locus is not determined by parent-of-origin markings; instead, it is defined by stochastic choice.

CTCF binds to its target motif at the *IGF2BP1* locus independently of DNA methylation

Binding of CTCF to its target motifs at both the human and mouse ICR of the *IGF2*/*H19* locus is sensitive to DNA methylation [10,26]. To test whether monoallelic expression of *IGF2BP1* in some individuals is also regulated by monoallelic DNA methylation of CTCF binding motifs, we examined a role for CpG methylation and allele-specific binding of CTCF at this locus.

To precisely determine the DNA sequence required for CTCF binding at the *IGF2BP1* locus, we searched for potential motifs using SOMBRERO [27], a *de novo* motif-finding algorithm that uses multiple self-organizing maps (SOM) to cluster sequences of a specific length (reads) from a set of input sequences (such as enriched genomic loci identified by ChIP-chip experiments). Motif alignment using STAMP [28] and comparison to the JASPAR transcription factor database [29] identified a distinct cohort of 68 motif models, all of which were identical to the canonical CTCF motif previously reported (Figure S5 in Additional file 2) [30]. The clustered reads associated with all 68 motif models

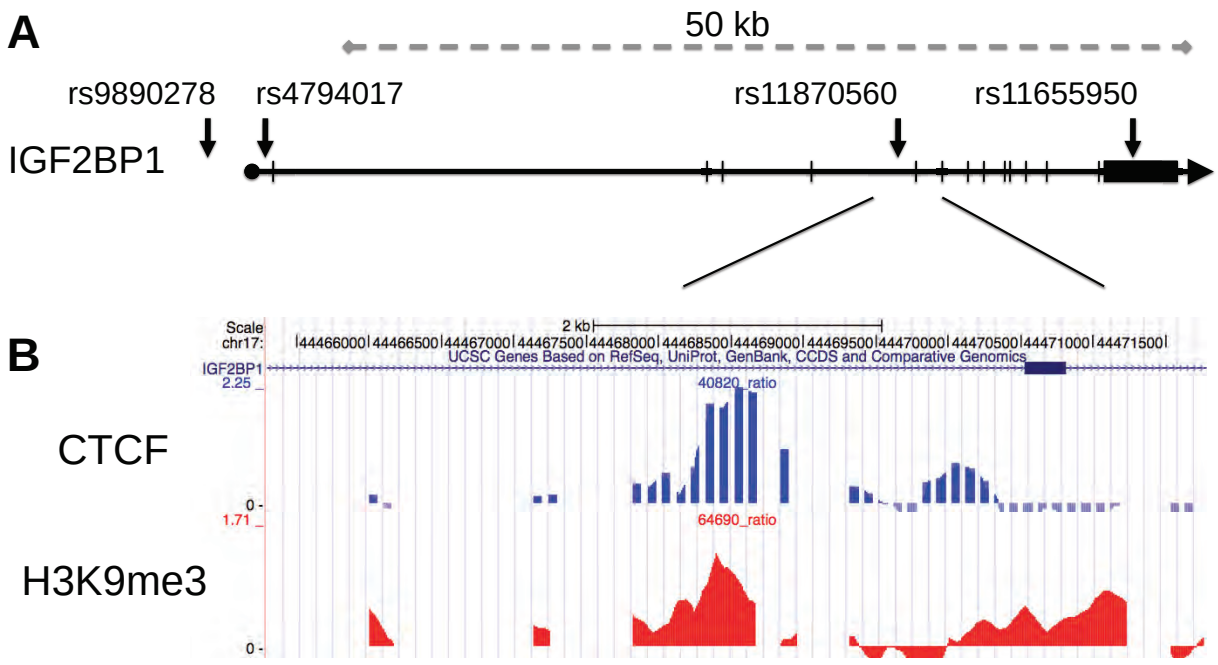


Figure 1 Colocalization of CTCF and H3K9me3 at the *IGF2BP1* locus. Array-based chromatin immunoprecipitation (ChIP-chip) data for both CTCF and histone H3 trimethylated at lysine 9 (H3K9me3) identify candidate loci for analysis of monoallelic expression. **(A)** Depiction of the *IGF2BP1* gene with specific SNPs examined in this study (arrows). **(B)** Close-up portion of the locus with tracks for CTCF enrichment (top track) and H3K9me3 association (bottom track) near SNP site rs11870560. The ChIP-chip data are displayed using the UCSC Genome Browser. DNA derived from CTCF ChIP experiments was analyzed by using microarrays with hybridization probes spaced 100 bp apart. The higher resolution of the H3K9me3 ChIP-chip data is due to the use of condensed array sets that tiled through all of the CTCF-positive regions with probes overlapping each other by 12 nt.

were mapped back to sequences enriched in our ChIP-chip analysis and were displayed using the UCSC Genome Browser (Figure S6 in Additional file 2). Using this approach, we identified 28,713 peaks, each composed of multiple overlapping reads, within the original 8,462 ChIP-chip loci. Using a strategy similar to that used to study ChIP-seq clustering [31], our frequency analysis of these peak heights yielded a bimodal distribution with an evident power law at low peak heights deviating to a clear excess in the numbers of peaks with heights > 10 (Figure S7 in Additional file 2). We consequently partitioned the peak populations into low-confidence and high-confidence groups using the peak height threshold of 10 (Figure S8 in Additional file 2).

Using this approach, we identified three potential motifs (X, Y and Z) (Figure 3) within the 350-bp region of the *IGF2BP1* gene locus enriched in our ChIP-chip experiments. Two of the putative binding sites, Y and Z, accumulated a significant number of matches to motif models. However, only one of the three putative CTCF binding sites belongs to the group of high-confidence binding sites (site Y) (Figure 3). In support of our *in silico* analysis of CTCF binding, previously published high-resolution ChIP-seq data on CTCF binding revealed enrichment of sequences surrounding motifs Y and Z (Figure 3A), suggesting that either one or both motifs is required for CTCF recruitment.

To further define the contribution of motifs Y and Z to CTCF binding, we measured their ability to recruit CTCF *in vitro* using immobilized template assays (Figures 3B and 3C). Wild-type and mutant DNA templates containing either one or both motifs were linked to magnetic beads, incubated with nuclear extract, washed and tested for association with CTCF by performing Western blot analysis. A 105-bp template containing the wild-type *IGF2BP1* intronic sequence efficiently recruited CTCF (Y^{wt} 105-bp template) (Figure 3B). In contrast, CTCF binding was severely reduced when the putative CTCF motif Y was mutated by four base substitutions (Figure 3C and Figure S9 in Additional file 2). To test the contribution of the adjacent motif Z to CTCF binding at the *IGF2BP1* locus, we generated several 125-bp DNA templates that encompassed both CTCF target motifs (Figure 3B). Targeted mutations at specific positions of motif Y and/or motif Z were introduced to test the contribution of each motif to recruitment of CTCF. Detailed sequences are shown in Figure S9 in Additional file 2. As shown in Figure 3C, the 125-bp template recruited CTCF more efficiently than the 105-bp template. However, motif Z does not contribute

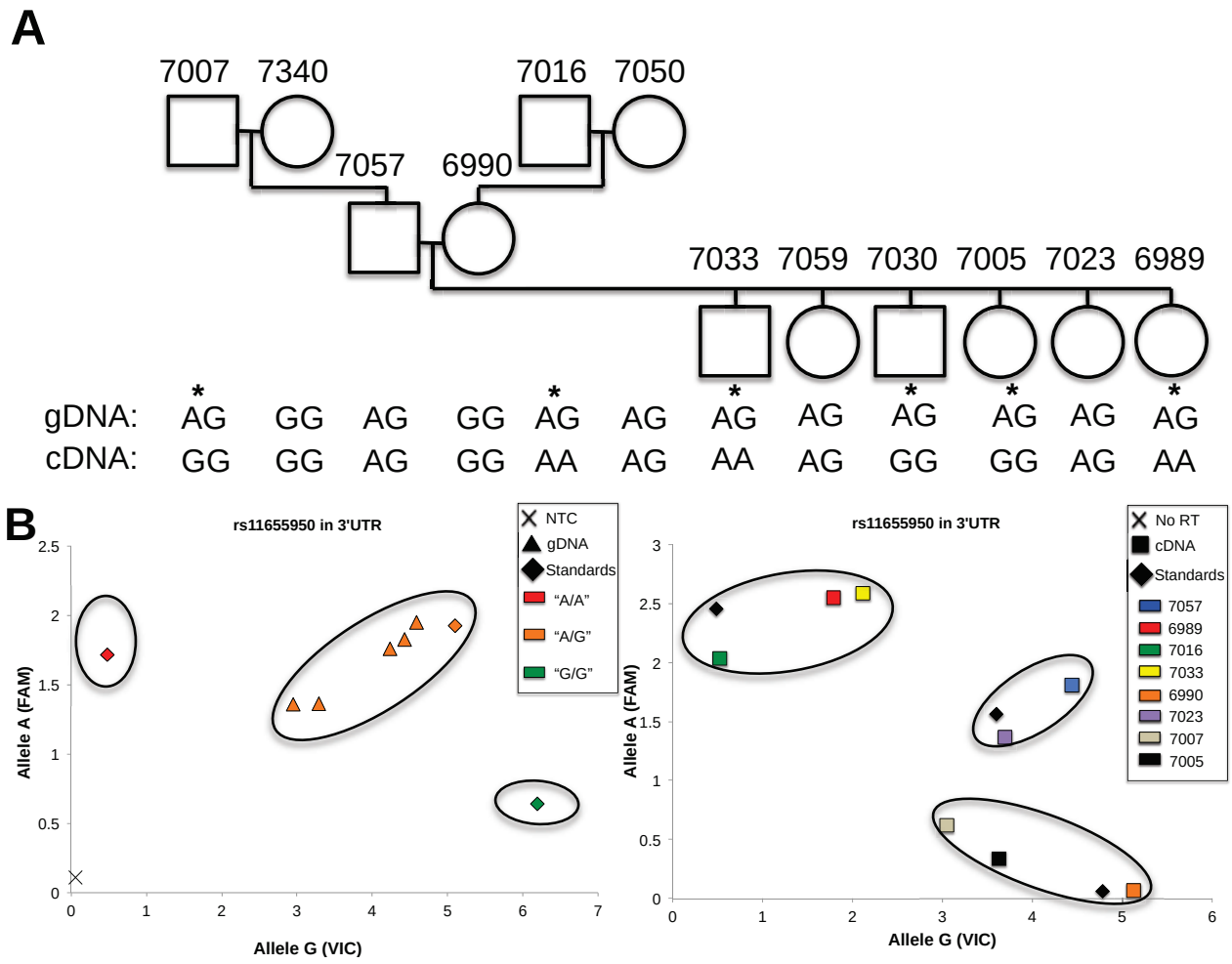


Figure 2 Analysis of allele-specific *IGF2BP1* expression. Comparative analysis of sequence variations in B lymphoblasts of the CEPH pedigree family 1331 reveals monoallelic expression of the *IGF2BP1* gene. **(A)** Pedigree analysis was carried out for the SNP site rs11655950 located in the 3'-UTR of the *IGF2BP1* gene. Each individual is shown with CEPH family identification, sample identification and genetic information (SNP genomic DNA (gDNA) genotype- or transcript-derived genotype). Individuals with monoallelic *IGF2BP1* gene expression are indicated by asterisks. If the individual is homozygous at the SNP, allele-specific expression cannot be defined. **(B)** Left: Genotyping results at rs11655950 with gDNA from members of CEPH family 1331. gDNA was analyzed using the TaqMan SNP Genotyping Assay. This assay discriminates between sequence variants using two allele-specific probes carrying two different fluorophores, VIC and FAM. Individuals coded in red and green represent cell lines that are homozygous for alleles A and G, respectively. Orange-labeled individuals contain both A and G alleles at SNP rs11655950 and represent informative cell lines used for further analysis of monoallelic expression. Diamonds indicate cDNA samples, and black × indicates averaged triplicates of a no-template control (NTC) near the origin of the graph. Right: Genotyping results of transcript-derived cDNA from heterozygous B lymphoblasts. Individuals are color-coded in the figure key. No-RT controls (No RT) from cDNA synthesis are shown near the origin of the graph and are indicated by a black X. Control samples (standards) of stem cell lines previously genotyped as homozygous AA, heterozygous AG and homozygous GG were plotted and are indicated by diamonds.

to CTCF recruitment, since targeted mutations in motif Z do not influence the level of CTCF binding. Consistent with this notion, CTCF binding is undetectable in the absence of a wild-type motif Y (Figure 3C).

CTCF binding site Y at the *IGF2BP1* gene contains a single CpG residue adjacent to the 14-bp core sequence of CTCF (Figure 4A). To establish whether binding of CTCF to Y^{wt} is inhibited by cytosine methylation, we tested Y^{wt} 105-bp immobilized templates after *in vitro* methylation of cytosine residues by CpG methyltransferase M.Sssl. For comparison, we examined CTCF motifs containing a higher CpG content, including site A of the MYC gene [32] as well as the B1 sequence of the ICR of the human *IGF2/H19* locus [10]. Cytosine methylation at the human B1 sequence is known to inhibit binding of CTCF. Consistent with this, recruitment of CTCF *in vitro* to immobilized templates containing the B1 sequence or the MYC site A is highly sensitive to DNA methylation (Figure 4B, top). In contrast, CpG methylation of the Y^{wt} motif has no effect on CTCF recruitment. Replacement of the Y^{wt} core

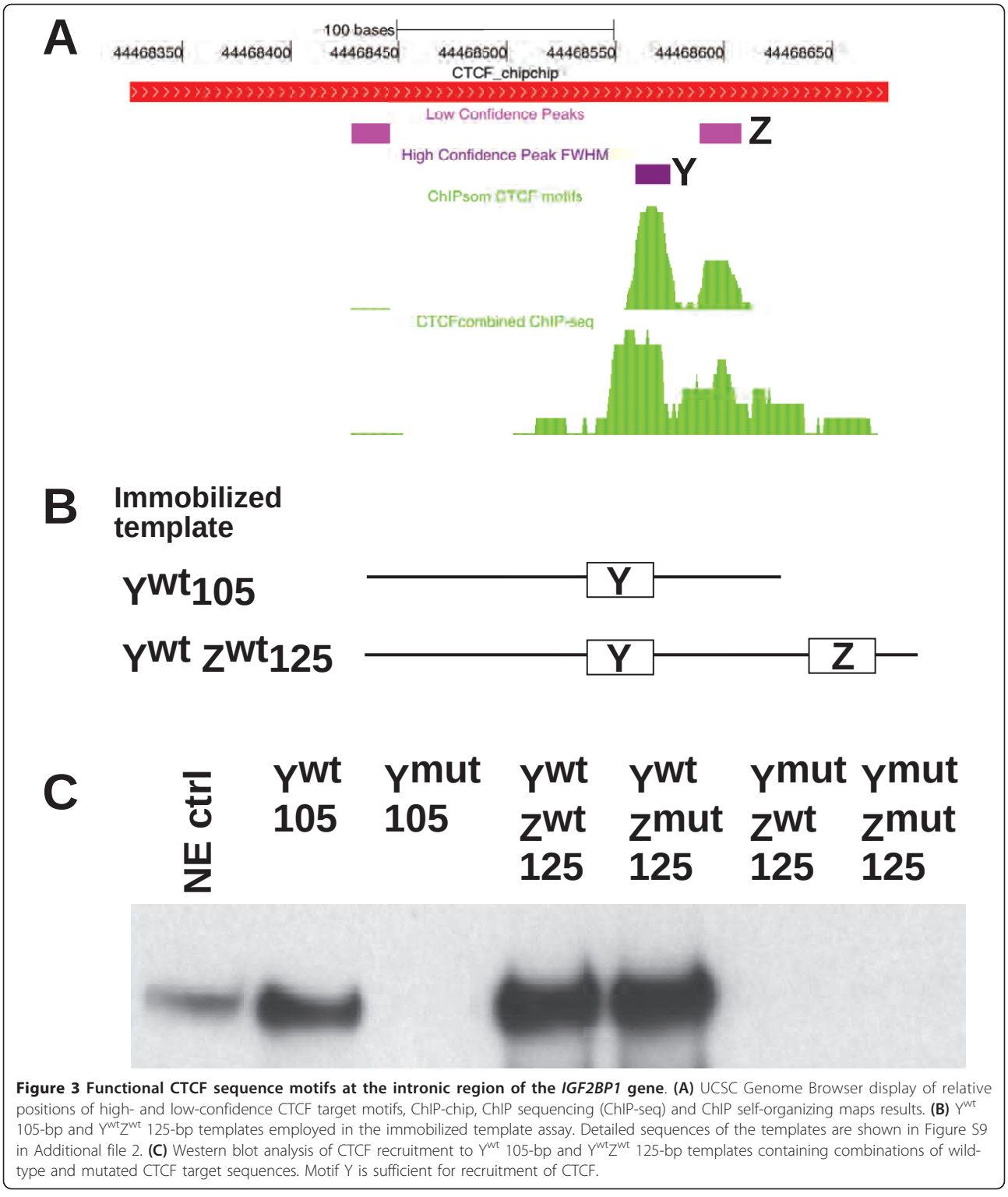


Figure 3 Functional CTCF sequence motifs at the intronic region of the *IGF2BP1* gene. **(A)** UCSC Genome Browser display of relative positions of high- and low-confidence CTCF target motifs, ChIP-chip, ChIP sequencing (ChIP-seq) and ChIP self-organizing maps results. **(B)** Y^{wt} 105-bp and $Y^{wt}Z^{wt}$ 125-bp templates employed in the immobilized template assay. Detailed sequences of the templates are shown in Figure S9 in Additional file 2. **(C)** Western blot analysis of CTCF recruitment to Y^{wt} 105-bp and $Y^{wt}Z^{wt}$ 125-bp templates containing combinations of wild-type and mutated CTCF target sequences. Motif Y is sufficient for recruitment of CTCF.

motif by the CTCF-binding sites of the chicken FII insulator element yields similar results. However, CTCF binding becomes sensitive to CpG methylation upon modification of the core motif to the mouse R3 sequence, a homologue of the human B1 sequence. In combination, despite the presence of a methylable CpG residue, binding of CTCF to the Y^{wt} sequence of the *IGF2BP1* gene *in vitro* is not sensitive to CpG methylation.

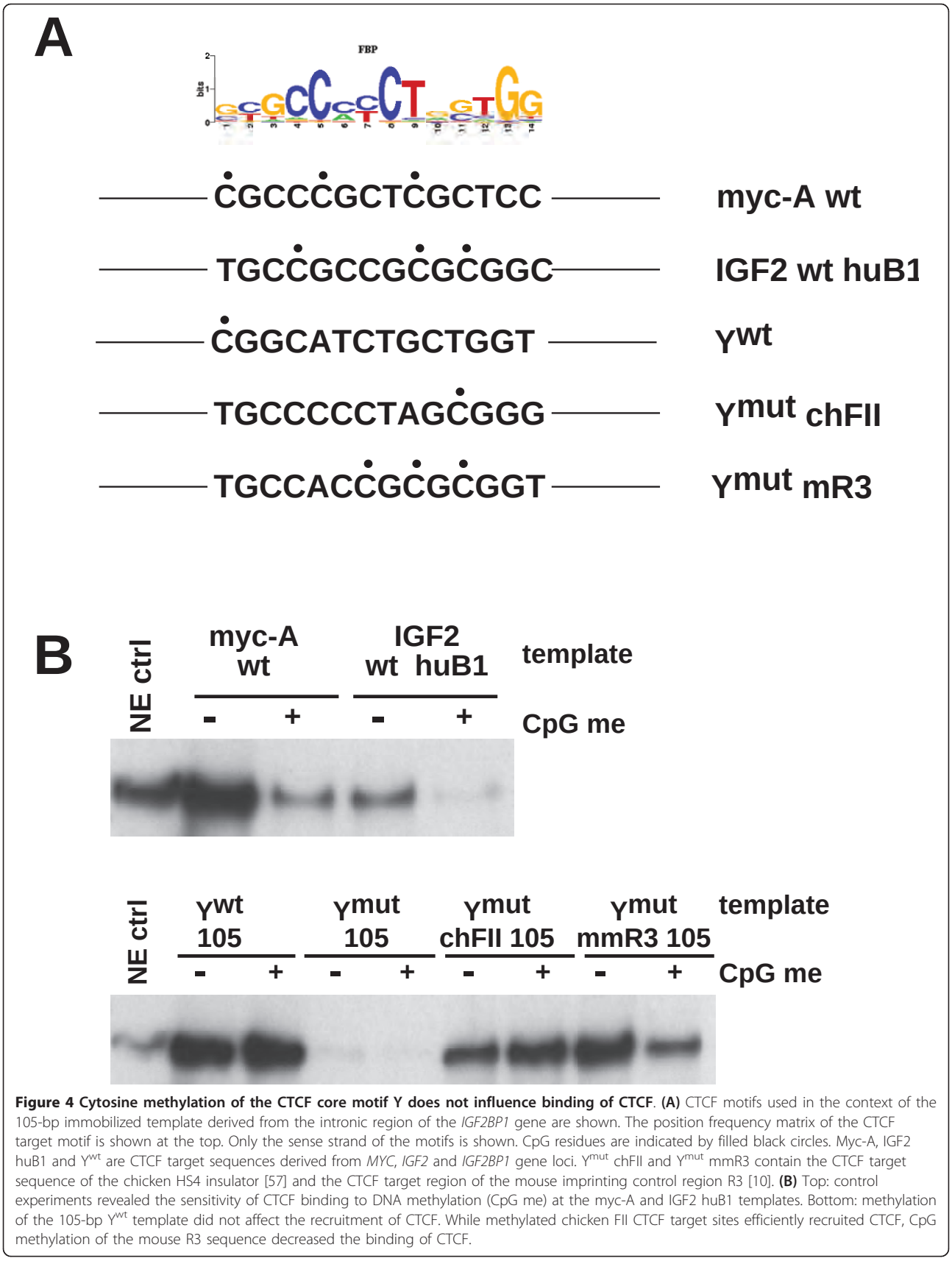


Figure 4 Cytosine methylation of the CTCF core motif Y does not influence binding of CTCF. **(A)** CTCF motifs used in the context of the 105-bp immobilized template derived from the intronic region of the *IGF2BP1* gene are shown. The position frequency matrix of the CTCF target motif is shown at the top. Only the sense strand of the motifs is shown. CpG residues are indicated by filled black circles. Myc-A, IGF2 huB1 and Y^{wt} are CTCF target sequences derived from *MYC*, *IGF2* and *IGF2BP1* gene loci. Y^{mut} chFII and Y^{mut} mmR3 contain the CTCF target sequence of the chicken HS4 insulator [57] and the CTCF target region of the mouse imprinting control region R3 [10]. **(B)** Top: control experiments revealed the sensitivity of CTCF binding to DNA methylation (CpG me) at the myc-A and IGF2 huB1 templates. Bottom: methylation of the 105-bp Y^{wt} template did not affect the recruitment of CTCF. While methylated chicken FII CTCF target sites efficiently recruited CTCF, CpG methylation of the mouse R3 sequence decreased the binding of CTCF.

To confirm that our *in vitro* characterization of CTCF binding accurately reflected the *in vivo* association of CTCF with the *IGF2BP1* locus, we evaluated the methylation status of the CTCF motif and adjacent CpG residues in the *IGF2BP1* intronic region in both biallelically (GM7057) and monoallelically (GM6989) expressing cells by using bisulfite sequencing (Figure 5). The methylation levels were calculated using BiQ Analyzer software [33]. Our data reveal that the CpG residue at the 5' end of the CTCF binding motif Y is invariably methylated. In addition, other methylable residues in this region exhibited some degree of DNA methylation. To further confirm binding of CTCF to methylated *IGF2BP1* intronic sequences, we bisulfite-sequenced DNA derived from immunoprecipitates of ChIP experiments with CTCF antibodies. As a control, we bisulfite-sequenced the *IGF2BP1* region derived from anti-H3K9me3 ChIP experiments. The results confirmed our *in vitro* finding that demonstrated an association of CTCF with a methylated motif (Figures 5B and 5C).

CTCF and H3K9me3 colocalize at both the maternal and paternal *IGF2BP1* alleles

Consistent methylation of the CTCF-binding motif in *IGF2BP1* indicated that DNA methylation is not allele-specific. To directly determine whether CTCF is bound monoallelically, we determined the allele-specific association of both CTCF and H3K9me3 by sequencing DNA recovered from ChIP experiments. We first identified informative cell lines by genotyping individuals from CEPH pedigree 1331 at SNP sites located close to the CTCF binding site. Cell lines derived from both monoallelically (GM7016 and GM6989) and biallelically (GM7057) expressing individuals were heterozygous at SNP site rs11870560 at the CTCF site (Figure 6A). We first applied the allelic discrimination assay to serial

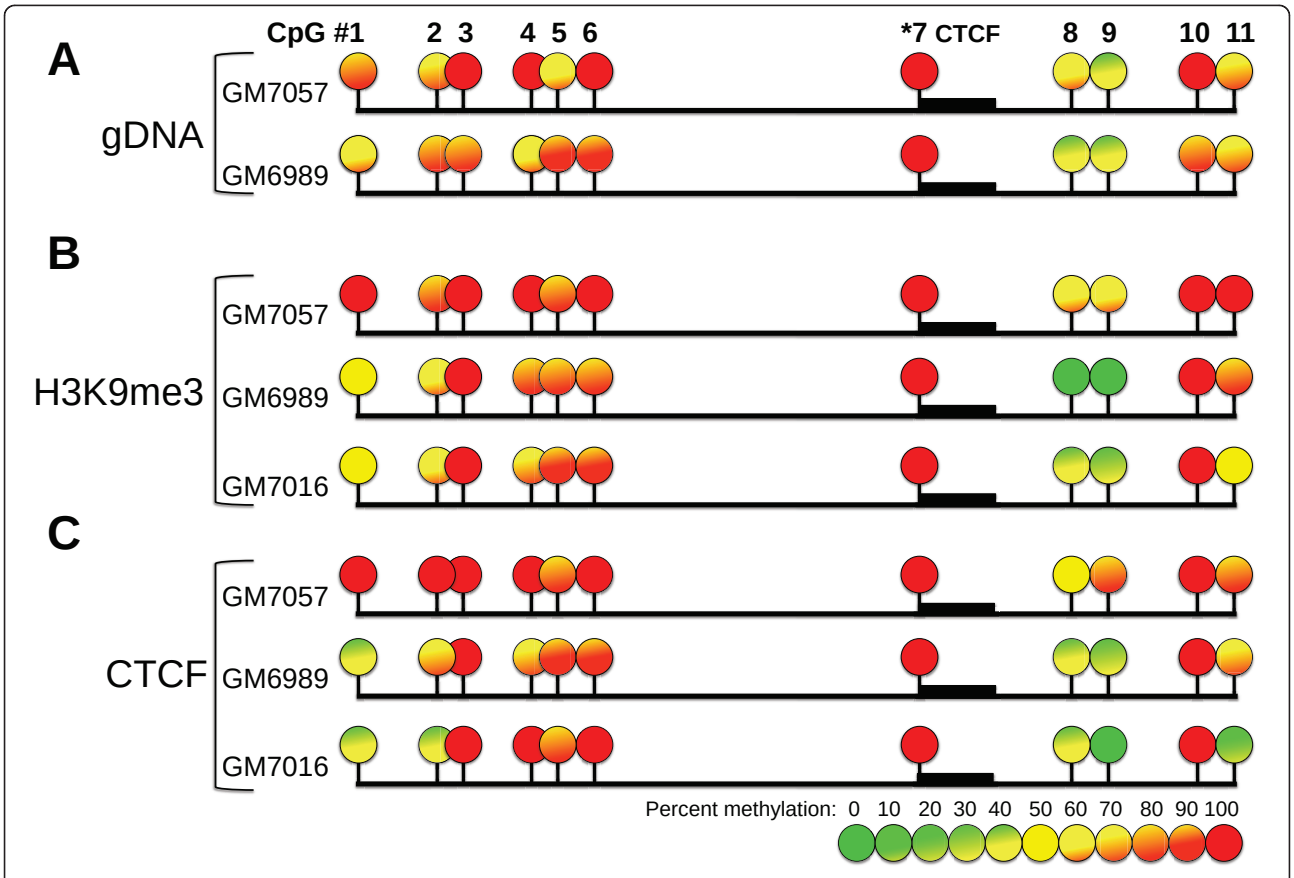


Figure 5 DNA methylation analysis of the *IGF2BP1* CTCF binding region. Analysis of DNA methylation with bisulfite sequencing at the intronic CTCF binding region of the *IGF2BP1* gene is shown. **(A)** The percentage of methylation of CpG sites in gDNA derived from cell lines that express *IGF2BP1* from only one allele (GM7016, GM6989) or from both alleles (GM7057) is shown. The CpG residue located within the CTCF binding motif is invariably methylated and is indicated by the thick black bar located adjacent to CpG site 7 (indicated by asterisks). **(B)** The percentage of methylation at each CpG site of the *IGF2BP1* CTCF site in DNA samples recovered from anti-H3K9me3 ChIP. **(C)** The percentage of methylation at each CpG site of the *IGF2BP1* CTCF site in DNA samples recovered from anti-CTCF ChIP experiments. The level of DNA methylation is represented according to the heat map keys located at the bottom of the figure.

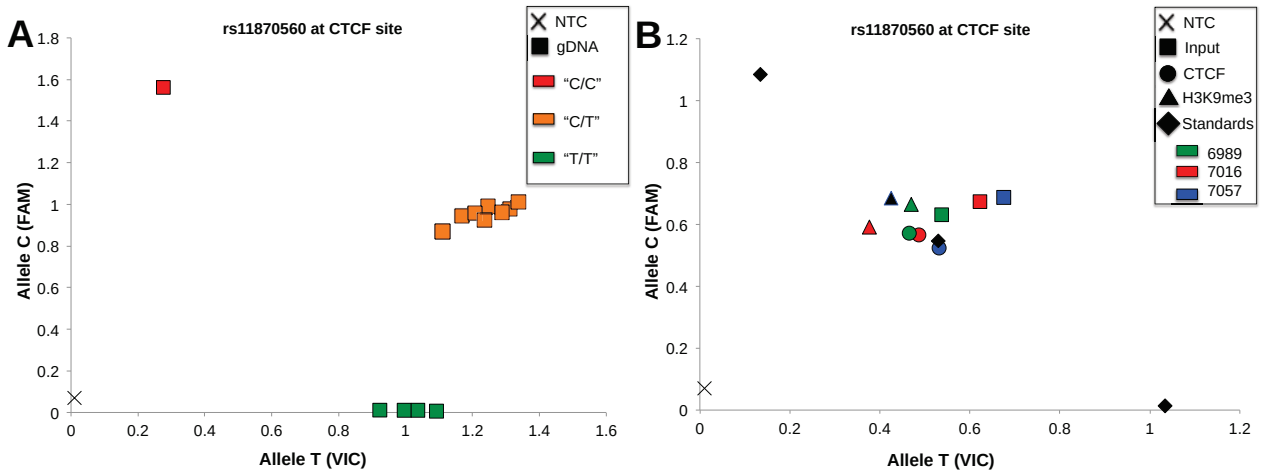


Figure 6 Allelic specificity of CTCF and H3K9me3. Informative ChIP templates were analyzed using the TaqMan allelic discrimination assay to address the allelic association of CTCF and H3K9me3. **(A)** Genotyping results at rs11870560 identify informative cell lines useful for the detection of allele-specific association of CTCF and H3K9me3. gDNA obtained from monoallelic and biallelic cell lines were genotyped using the TaqMan allelic discrimination assay. Squares represent gDNA samples and are coded in red and green to represent cell lines that are homozygous for allele C and allele T, respectively. Orange indicates heterozygous individuals. Averaged triplicate of a no-template control (NTC) is shown near the origin of the graph. **(B)** Genotyping at SNP rs11870560 with DNA templates recovered from ChIP experiments was used to identify the enrichment of the two alleles with either CTCF (circle) or H3K9me3 (triangle). Each color shown in the figure key represents a lymphoblastoid cell line (LCL) derived from an individual of the pedigree, while the shape represents the source of each sample (for example, squares signify input samples, while circles and triangles indicate ChIP samples obtained with CTCF and H3K9me3 antibodies, respectively). Immunoprecipitated templates were generated using the ChIP protocol described in Materials and Methods. Both monoallelic and biallelic cell lines indicate biallelic distribution of both CTCF and H3K9me3. Diamonds indicate control LCL samples (standards) previously genotyped as homozygous CC, heterozygous CT and homozygous TT.

dilutions of known homozygotes of the two possible alleles to test its ability to quantitatively assess the contribution of each allele in a DNA mixture. This assay provides quantitative results with high sensitivity and reproducibility within a ten-fold range of DNA concentrations, thus making it a useful tool for allelic discrimination of immunoprecipitated DNA (Figure S4 in Additional file 2). We used two monoallelically (GM7016 and GM6989) and one biallelically (GM7057) expressing cell lines to genotype DNA recovered from ChIP assays using either anti-CTCF or anti-H3K9me3 antibodies. Each analysis was performed in triplicate. Equal proportions of the two sequence variants were detected in DNA derived from ChIP assays with either H3K9me3 or CTCF antibodies, indicating that CTCF associates with both the maternal and paternal alleles (Figure 6B). Thus, monoallelic expression of the *IGF2BP1* gene is not mediated through monoallelic binding of CTCF.

The *IGF2BP1* promoter associates with both active and silent histone modifications in B cells

To define alternative mechanisms responsible for random monoallelic expression of *IGF2BP1*, we sought to identify markers that distinguish the active and inactive alleles. K27-trimethylated and K4-trimethylated histone H3, respectively, mark transcriptionally silent and active chromatin. We determined the relative enrichment of these two histone markers at the *IGF2BP1* promoter for each allele in both monoallelically and biallelically expressing cell lines using ChIP with anti-H3K4me3 and anti-H3K27me3 antibodies. Both H3K4me3 and H3K27me3 were detected at the *IGF2BP1* gene promoter (Figure 7A). To determine whether any of the histone modifications selectively associates with either allele, we again searched for informative sequence SNPs at the *IGF2BP1* promoter region in the CEPH pedigree. Cell lines derived from individuals GM6989 (monoallelically expressing cell line) and 7057 (biallelically expressing cell line) were heterozygous at SNP rs9890278 located upstream of the transcription initiation site, whereas GM7007 (monoallelically expressing cell line) was heterozygous for SNP rs4794017 located 1 kb downstream of the transcription initiation site. To address whether active and silent alleles in these cell lines are distinguished by specific histone markers, we sequenced SNPs rs9890278 and rs4794017 in gDNA recovered from ChIP experiments using anti-H3K4me3 and anti-H3K27me3 antibodies. The results revealed that both H3K4me3 and H3K27me3 are detected on both alleles in a bivalent fashion (Figure 7). In combination, our results indicate that both active and silent

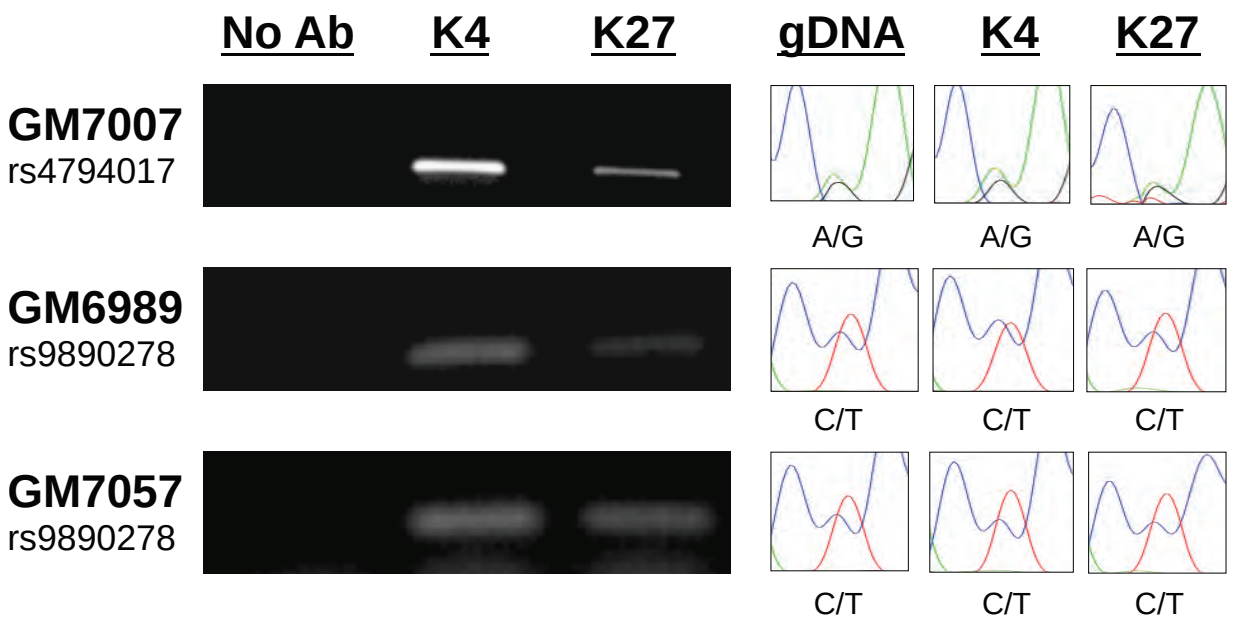


Figure 7 *IGF2BP1* promoter region is enriched with activating and silencing chromatin modifications. DNA recovered from ChIP experiments using anti-H3K4me3, anti-H3K27me3 and RNA polymerase II ChIP templates was genotyped by sequencing the *IGF2BP1* promoter region containing sequence variant rs4794017 or rs9890278. Left: Enrichment of H3K4me3 (K4) and H3K27me3 (K27) in monoallelically (GM7007, GM6989), and biallelically (GM7057) expressing cell lines. The positions of informative SNPs rs479017 and rs9890278 are shown in Figure 1. Both activating and silencing marks are significantly enriched. Right: Sequences enriched by ChIP were excised and sequenced. The results show an association of both alleles with active and silent histone modifications at the *IGF2BP1* promoter region independent of transcriptional status.

histone markers (H3K4me3 and H3K27me3) coexist in the promoter region of both *IGF2BP1* alleles in monoallelically as well as biallelically expressing cell lines. These data indicate that allele-specific expression of *IGF2BP1* cannot be explained by differential association of active and silent histone markers.

Silencing of the inactive *IGF2BP1* allele by inhibition of RNA polymerase II elongation

Monoallelic expression of *IGF2BP1* cannot be attributed solely to selective activation or silencing of one allele through histone modifications, since H3K4me3 as well as H3K27me3 are detected at both alleles. H3K4me3 is typically associated with transcriptionally active alleles, raising the question whether allele-specific transcription elongation or RNA processing accounts for monoallelic expression of the *IGF2BP1* gene. To address this hypothesis, we again searched mono- and biallelically expressing cell lines for sequence SNPs near the site of transcription initiation at the *IGF2BP1* promoter. Within CEPH pedigree 1331, only line GM7007 contained a heterozygous genotype at SNP site rs4794017 located within intron 1, 1 kb downstream of the transcription initiation site. We performed RNA polymerase II ChIP on chromatin prepared from this monoallelically expressing line. Quantitative real-time PCR analyses revealed enrichment of *IGF2BP1* promoter sequences similar to the enrichment observed at the MYC promoter. Immunoprecipitated DNA was PCR-amplified and sequenced (Figure 8A). Identification of both sequence variants at rs4794017 in DNA recovered from ChIP experiments indicates that RNA polymerase II is associated with both *IGF2BP1* alleles, which is consistent with the presence of H3K4me3 at the promoter of both alleles.

These data suggest that allele specificity of transcription is achieved after recruitment of RNA polymerase to both alleles, such as through transcriptional pausing and/or selective RNA processing. A major rate-limiting step in transcription elongation is pausing of RNA polymerase II in the promoter proximal region immediately downstream of the transcription initiation site [34-37]. We sequenced the 5' portion of the *IGF2BP1* gene of all monoallelically expressing cell lines to identify sequence variants that would be useful for allelic identification of promoter proximal regions occupied by RNA polymerase II or for the determination of the allelic origin of unspliced, precursor pre-mRNA transcripts. Since no additional informative sequence variants were identified, we focused on the detection and sequencing of pre-mRNA transcripts about 1 kb downstream of the transcription initiation site in GM7007. Using the informative SNPs located within intron 1 of this gene, we targeted nascent unspliced RNA with primers designed

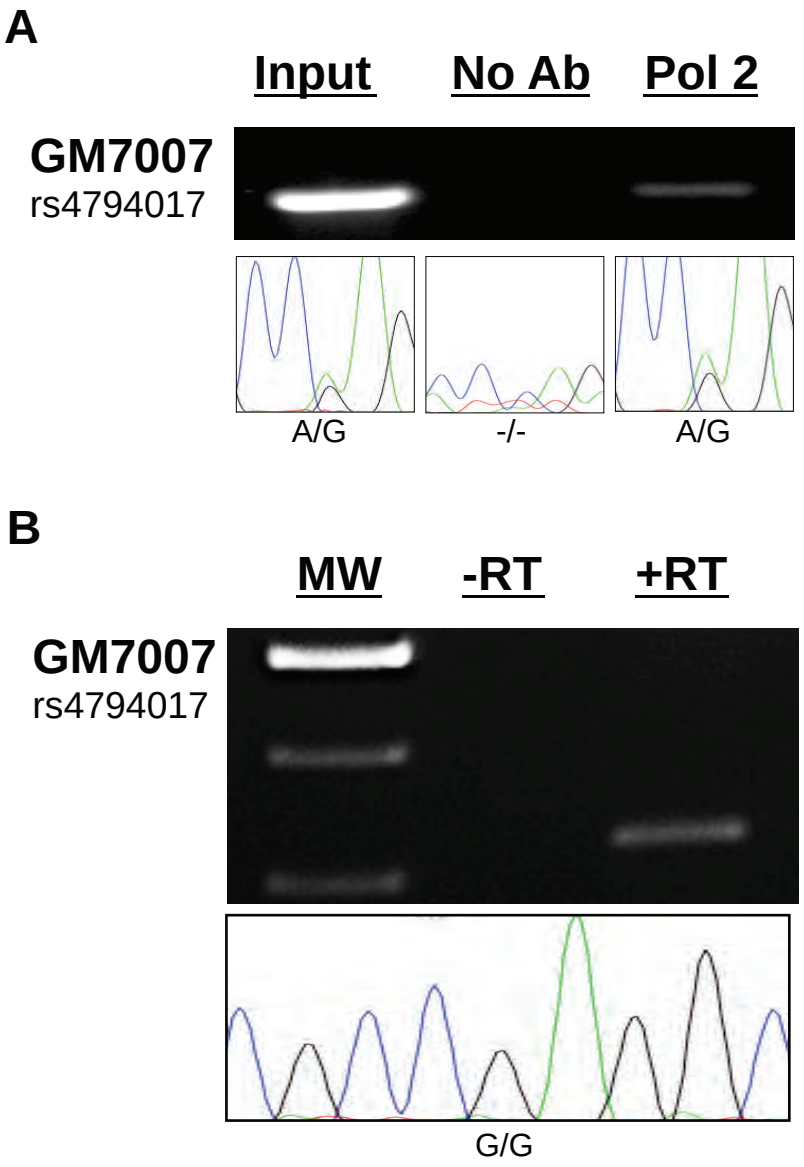


Figure 8 RNA polymerase II associates with both alleles in a monoallelically expressing cell line. **(A)** Recruitment of RNA polymerase II to the *IGF2BP1* promoter was examined by ChIP in monoallelically expressing GM7007 cells. DNA recovered from chromatin that had been immunoprecipitated with anti-RNA polymerase II antibodies (Pol2) was amplified and sequenced for allelic association. Sequencing results (bottom) reveal that both alleles of the monoallelically expressing cell line GM7007 associate with RNA polymerase II near SNP site rs4794017. In contrast, sequencing of DNA from "no antibody" ChIP reactions failed to produce sequence reads. **(B)** Allele specificity of precursor mRNA was determined by sequencing of cDNA prepared from total RNA of GM7007 cells. RNA had been extensively pretreated with DNase I to eliminate gDNA prior to reverse transcription by RT. Subsequently, cDNA samples were amplified using primers flanking rs4794017. In the absence of RT (-RT), no amplification products were oberved. +RT amplicons were gel-purified and sequenced. Bottom: Sequence traces at the heterozygous SNP site rs4794017 located 1 kb downstream of the transcription initiation site in cDNA of GM7007 indicate a single allele.

to amplify a region containing SNP site rs4794017. To avoid detection of gDNA in RNA samples, DNA was efficiently removed by treatment with an engineered, highly active form of DNase I (TURBO DNase I; Applied Biosystems/Ambion, Austin, TX, USA). This protocol allowed detection of pre-mRNA free of gDNA contamination (Figure 8B). Sequencing of amplified *IGF2BP1* pre-cDNA revealed only one of the two sequence variants at SNP rs4794017, indicating that pre-mRNA transcripts are transcribed from only one allele despite the presence of RNA polymerase II on both alleles. Thus, our data indicate that monoallelic expression of the *IGF2BP1* gene is regulated through allele-specific transcriptional elongation prior to SNP site

rs4794017, located approximately 600 bp downstream of the first intron splice site.

Discussion

Allele-specific expression in which one parental allele is stochastically or parent-of-origin-specifically silenced is widespread in mammalian organisms. Large-scale, allele-specific gene expression analyses have revealed that 5% to 10% of autosomal genes show random monoallelic transcription [7]. The stability of allele-specific expression through many cell passages suggests that epigenetic modifications maintain this specific type of gene regulation throughout generations of cells. Analogously to the regulation at the imprinted *IGF2*/*H19* locus, we tested the hypothesis whether monoallelic binding of CTCF, a characteristic marker for the *IGF2*/*H19* ICR, also underlies random monoallelic expression. Using ChIP-chip analyses, we identified chromosomal loci that are enriched in both CTCF and H3K9me3 and cross-correlated their positions with previously published lists of monoallelically expressed genes. Our data indicate that genomic loci enriched for both CTCF and H3K9me3 do not significantly correlate with monoallelically expressed genes. While this lack of correlation could be formally attributed to variations in monoallelic expression between different cell lines and types, it should be noted that the genome-wide pattern of CTCF binding is very consistent between different cell lineages [30,38,39]. Thus, if CTCF and H3K9me3 contribute to allele-specific expression, it should be detectable through allele-specific association of CTCF and H3K9me3. Focusing on the *IGF2BP1* gene, we tested whether monoallelic expression in a pedigree of LCLs correlates with monoallelic binding of CTCF. Although binding of CTCF to its targets is thought to be sensitive to DNA methylation, we surprisingly found the cytosine residue closely flanking the CTCF target motif at the *IGF2BP1* gene to be consistently methylated without any effect on CTCF recruitment. Indeed, our *in vitro* analyses of the binding requirements using immobilized templates confirmed that methylation of cytosine residues within the *IGF2BP1* sequence does not affect CTCF binding. These data are consistent with those in previous studies in which researchers found that cytosine methylation outside the CTCF core motif did not affect the binding affinity of bacterially expressed wild-type and mutant CTCF proteins [40]. This information is useful for the identification of the genomic subset of CTCF sites that might contribute to differential cell- and stage-specific expression due to their sensitivity to cytosine methylation, potentially mediating changes in large-scale chromatin organization during development and disease.

A number of studies have examined the correlation of allele-specific expression with allele-specific association of epigenetic markers [21,41-45]. The data produced by these studies have established common signatures of imprinted alleles, including H3K9me3 and H3K4me3, providing a powerful means by which to identify novel imprinted or monoallelically expressed loci [46-48]. In contrast to the strict allele-specific association of DNA methylation and chromatin markers at imprinted genes, histone modifications at the nonimprinted, monoallelically expressed *IGF2BP1* gene do not predict the active allele. Both H3K4me3 and H3K27me3, markers characteristic of active and inactive loci, are associated with each allele, as both sequence variants of SNP rs4794017 are present in the DNA of heterozygous individuals recovered from ChIP experiments. Moreover, loading of RNA polymerase II also does not provide a reliable marker for identifying the transcribed allele. Our ChIP experiments identified both sequence variants at SNP rs4794017 within the promoter proximal region of anti-RNA polymerase II immunoprecipitated DNA. Because only one LCL in our study was informative for determining an association of RNA polymerase II at the *IGF2BP1* alleles, we could not define how frequently this type of regulation occurs within cell lineages and throughout the genome. However, other investigators have reported similar results at the *PCNA* gene. Maynard *et al*. [44] found that both *PCNA* alleles in IMR90 cells are bound by RNA polymerase II, although only one allele generates full-length mRNA. Together, these data suggest that transcription elongation not only is a general rate-limiting step in the transcription of the vast majority of genes [34,35,37] but also regulates the expression of a subset of monoallelically expressed genes.

The expression of *IGF2BP1* in differentiated cell types, including LCLs, is significantly lower than in embryonic stem cells. In an attempt to determine whether allele-specific expression also contributes to *IGF2BP1* regulation early in development, we genotyped both gDNA and cDNA in 11 human embryonic stem cell (hESC) lines. However, while only three hESC lines were informative (heterozygous at SNP rs11655950), all three expressed *IGF2BP1* in a biallelic manner. Although the number of available and informative hESC lines is not sufficient to clearly define a role for allele-specific elongation in early developmental stages, we believe that it is unlikely that this mechanism is restricted to cell types with low levels of *IGF2BP1* expression. Control of transcriptional activity through promoter proximal pausing or premature termination of transcription is not restricted to specific gene classes characterized by low levels of transcriptional activity [35]. We speculate that distinct positioning of the homologous alleles within the nuclear space and association with distinct "transcription factories" may contribute to monoallelic transcription elongation.

The *IGF2BP1* gene is highly expressed during embryonic development and is required for the regulation of mRNA stability of several genes involved in growth regulation, including the *IGF2*, β-catenin and *MYC* genes [23-25]. Consistent with its role in early developmental stages, the *IGF2BP1* gene is downregulated in differentiated cell types, and overexpression of *IGF2BP1* is known to occur in multiple human cancers, including breast, lung and colon [49-52]. Thus, changes in the level of *IGF2BP1* expression through silencing of only one allele could provide a safeguard against pathogenesis and disease.

Conclusions

Allele-specific gene expression is common in the human genome and is thought to contribute to phenotypic variation. The allele-specific association of CTCF, H3K9me3 and DNA methylation is a characteristic marker of imprinted gene expression at the *IGF2*/*H19* locus, raising the question whether these epigenetic markers are useful for identifying both imprinted and random monoallelically expressed genes throughout the genome. In this study, we have demonstrated that colocalization of CTCF and H3K9me3 does not represent a reliable chromatin signature indicative of monoallelic expression. In addition, we conclude that allele-specific binding of CTCF requires methylation of very specific cytosine residues within the target motif, effectively limiting the number of CTCF binding sites potentially affected by allele-specific binding. In addition, the active and inactive alleles of random monoallelically expressed genes do not necessarily correlate with active or inactive histone markers. Remarkably, the selection of individual alleles for expression at the *IGF2BP1* locus occurs during early stages of transcription elongation.

Methods

ChIP-chip analyses

The amplification and preparation of immunoprecipitated DNA derived from HBL100 cells for hybridization to ENCODE arrays (Roche NimbleGen Inc., Madison, WI, USA) was performed essentially as described previously [53]. Sample labeling and array hybridization were performed at NimbleGen Systems Inc. Genomic control DNA was labeled with Cy3, and sample DNA was labeled with Cy5. Both Cy3- and Cy5-labeled DNA were hybridized to high-density arrays tiling through ENCODE regions with 50-mer oligonucleotides across nonrepetitive genomic regions. The ratios of the Cy3 and Cy5 intensities of each probe were calculated using NimbleGen Systems' proprietary software.

Peak detection and false-positive rate calculation

A genomic sequence was considered a possible CTCF-binding site if there were at least four probes among the sequence probe and the flanking probes within a window covering 250 bp on both sides of the probe had $\log_2$ ratio values above a specified cutoff value. The cutoff value was calculated separately for each chromosome. The cutoff value is a given percentage of the value (mean + 6 × standard deviation) of the $\log_2$ ratio values of all the probes covering the chromosome. The possible binding sites thus detected are called peaks. To calculate the false-positive rate (FPR) by data permutation, the $\log_2$ ratio values among probes were scrambled to generate a randomized data set for each individual chromosome. Multiple repetitions of this process generated 20 randomized data sets for each chromosome. Subsequently, the peak detection algorithm described above was applied to count the average number of peaks in the 20 randomized data sets using the same cutoff. The ratio of that number to the number of peaks from the nonrandomized data set is the FPR. The FPR is associated with the threshold setting, which is indicated by the value of cutoff *P*. Peak detection and randomization of data sets were repeated for different threshold settings of *P*. The corresponding FPRs were calculated and assigned to peaks. The FPR value assigned to the individual peaks is the value associated with the cutoff *P* at which the peak is first detected.

Peak discovery was performed using chromatin immunoprecipitate:input ratios combined from adjacent oligonucleotides within 250-bp regions. The FPR of detection was estimated by permutation analyses in which the experimentally determined $\log_2$ ratio values were reassigned to probes in a random fashion, allowing selection of stringency and specificity levels. To define sites of CTCF interaction with high confidence, peaks were required to be present in all three biological replicates and to be generated at a FPR < 0.05.

Chromatin immunoprecipitation

Chromatin was prepared for immunoprecipitation as described previously [54] by cross-linking the cells in 1% formaldehyde for 5 minutes and subjecting them to subsequent sonication until the bulk of DNA was 300 to 600 bp in size. Chromatin corresponding to 2×10^7 cells was immunoprecipitated with anti-CTCF antibody (D31H2; Cell Signaling Technology, Danvers, MA, USA), anti-H3K9me3 antibody (ab8898; Abcam, Cambridge, MA, USA), anti-trimethyl K4-histone H3 antibody (ab8580; Abcam), anti-trimethyl K27-histone H3 antibody (Millipore 07-449, Billerica MA, USA) or anti-RNA polymerase II antibody (sc899; Santa Cruz Biotechnology, Santa Cruz, CA, USA). Immunoprecipitates were washed, the DNA protein cross-links were reversed and the recovered DNA was tested by performing conventional quantitative PCR as described previously [54]. RNA polymerase II ChIP experiments were performed

using the Matrix ChIP protocol [55]. Sequences of primers specific for the gene loci under study as well as the reference primers are available upon request.

RNA extraction and RT-PCR

Synthesis of cDNA was carried out according to the manufacturer's instructions (Qiagen, Valencia, CA, USA) using 1 μg of total RNA. For detection of pre-mRNA, RNA preparations were pretreated with TURBO DNase I (Ambion/Applied Biosystems) as described in the manufacturer's protocol. RT was carried out at 37°C for one hour.

Cell culture

Cell lines were cultured in RPMI 1640 medium supplemented with 10% FCS, 2 mM L-glutamine and the antibiotics penicillin (50 U/mL) and streptomycin.

Sodium bisulfite conversions

gDNA was treated with sodium bisulfite using the EZ DNA Methylation Kit (Zymo Research, Orange, CA, USA) according to the manufacturer's instructions. PCR amplification of bisulfite-treated DNA was performed using ZymoTaq DNA Polymerase (Zymo Research Corporation, Irvine, CA, USA) and conversion-specific primers targeted to the *IGF2BP1* CTCF region (forward primer: 5'-TATTTTTTAGTTGGGTTAAT-TGGTG-3', reverse primer: 5'-ATACTACCTCTCCTTCCAAA ATCTC-3'). The amplified products were purified by gel electrophoresis and sequenced. Each case was scored as methylated or unmethylated, and the percentage of methylation was calculated using BiQ Analyzer software [33].

TaqMan allelic discrimination assays

TaqMan allelic discrimination assays were performed according to the manufacturer's instructions with the following adjustments: cDNA from B lymphoblasts was preamplified for 14 cycles. PCR products were gel-purified and subsequently used as templates in the genotyping of samples. The specific primer sequences used are avaliable upon request.

In vitro CTCF binding analysis using immobilized templates

Crude nuclear extract was prepared from 1×10^9 Jurkat cells grown in growth media (RPMI 1640 with 10% fetal bovine serum) according to methods described previously [56]. Biotinylated template DNA was generated by PCR amplification of the *IGF2BP1* intronic region using a biotinylated/nonbiotinylated primer combination. The specific primer sequences are available upon request. For each binding reaction, 1 pM biotinylated DNA template was coupled to 50-μg streptavidin-linked magnetic beads (Dynabeads M-280 Streptavidin; Invitrogen, Carlsbad, CA, USA). Templates immobilized to magnetic beads were washed three times in B&W buffer (5 mM Tris, pH 7.5, 0.5 mM ethylenediaminetetraacetic acid (EDTA), 1 M NaCl) and resuspended in Jurkat nuclear extract. After a two-hour incubation at 4°C, immobilized templates were washed three times in Dignam buffer D (20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid, pH 7.9, 20% glycerol, 0.1 M KCl, 1 mM EDTA, 0.1 mM ethylene glycol tetraacetic acid, 1% Nonidet P-40, 1 mM dithiothreitol) containing protease inhibitor (P8340; Sigma, St Louis, MO, USA). To recover template-bound proteins, beads were incubated in elution buffer (5 mM Tris, pH 7.5, 0.5 mM EDTA, 1 M $NaHCO_3$) including protease inhibitors. After a 5-minute incubation, the eluate was removed and transferred into a fresh tube. The presence of CTCF in the eluate was determined using standard Western blot analysis protocols.

Additional material

Additional file 1: Table S1. Genomic coordinates of 293 genomic sites that are marked by both CTCF and H2K9me3. Table S2. List of genes tested for monoallelic expression in lymphoblastoid cell lines.

Additional file 2: Figure S1. Detection and colocalization of CTCF and H3K9me3 at the human IGF2-H19 ICR locus by ChIP-chip experiments. Top: Enrichment of CTCF binding sites. Middle: Results of large-scale array-based chromatin immunoprecipitation (ChIP-chip) survey of histone H3 trimethylated at lysine 9 (H3K9me3) binding. Bottom: H19 exons demonstrating positions of CTCF binding and histone modifications relative to exons. **Figure S2. Analysis of the clonal status of lymphoblastoid cell lines used in this study**. Following the protocol described in [22], PCR amplification of two regions within the variable segment in the immunoglobulin heavy chain gene (conserved framework region 2 (Fr2) and the variable joining regions (VLJH)) reveals the clonal status of lymphoblastoid cell lines (LCLs). The amplification product from a polyclonal population (P) gives rise to fragments of varying length due to the large number of rearranged immunoglobulin genes and appears as a broad band. Amplification of DNA derived from monoclonal cell lines results in one or two discrete bands within an expected size range of 240 to 280 bp. The polyclonal sample (P) was obtained from the peripheral blood of a healthy donor. Lanes 1 through 4: monoclonal cell lines GM7007, GM7033, GM6989 and GM7030. Lanes 5 through 8: monoclonal lines GM7050, GM7023, GM7059 and GM7057. MW, DNA size marker. **Figure S3. Sequencing results give results identical to those derived from the TaqMan allelic discrimination assay**. **(A)** Standard sequencing results of two individuals at SNP site rs9904288. **(B)** TaqMan allelic discrimination assay confirms the heterozygosity of GM7057 and the homozygosity of GM6990. **Figure S4. Quantitative assessment of TaqMan genotyping using specific probe set at SNP rs11655950**. The 3'-UTR of the *IGF2BP1* gene was amplified using primers given in Supplemental Table 2. This segment contains an A/G SNP. The PCRs included a FAM-labeled probe for the A allele and a VIC-labeled probe for the B allele. After PCR amplification, an end point fluorescence reading was taken on the ABI PRISM 7700 with SDS version 1.4 software (Applied Biosystems). The determination of the quantitative assignment of known genotypes is plotted. Concentration dilutions were created using known homozygous cell lines. Preparations of gDNA samples shown represent the following allele B/allele A ratios: 100:0, 80:20, 60:40, 50:50, 40:20, 20:80 and 0:100. Heterozygosity was based on the fluorescence intensity of FAM, VIC or both dyes together. Error bars indicate 5% of triplicate sample value. Allele A curve yields $y =$

0.0102x + 0.0415 with R^2 = 0.98934. Allele B curve yields y = -0.0085x + 0.9796 with R^2 = 0.98196. **Figure S5. Phylogenetic tree of motifs determined from motif analysis of the 8,462 loci derived from the ChIP-chip analysis using STAMP**. All members of the highlighted group have matches identical to the canonical CTCF motif model as part of the JASPAR transcription factor binding site database. The resulting familial binding profile for all 68 such models is displayed. **Figure S6. Fine mapping of CTCF motifs in sequences enriched in ChIP-chip experiments**. Motif reads were mapped onto the genomic loci defined by ChIP-chip for CTCF binding. The extent of the ChIP-enriched sequences is indicated by red bar. Several read clusters are apparent and vary in depth and spatial extent (green areas). **Figure S7. Frequency distribution of cluster depth for all motif clusters**. A power law is apparent for clusters of depth ≤ 10 with evident deviation in the population and a maximum of about 40. The vertical green line demarcates the low and high confidence clusters. **Figure S8. Discrimination between high- and low-confidence sites**. The region shown in Supplemental Figure S6 is annotated by overlaying enriched sequences with high- and low-confidence tracks. **Figure S9. Sequences of immobilized templates used in *in vitro* binding experiments**. CTCF core motifs Y and Z are underlined. Site-specific mutations in either the Y or Z motif are highlighted in yellow. In Y^{mut} chFII and Y^{mut} mmR3, site-specific mutations (highlighted in green) were introduced to generate CTCF motifs identical to the chicken HS4 FII site and the mouse imprinting control region R3. The *IGF2* wild-type huB1 sequence is derived from the human *IGF2* imprinting control region containing the methylation-sensitive CTCF binding site B1.

Abbreviations

FCS: fetal calf serum; PCR: polymerase chain reaction; RT: reverse transcriptase; SNP: single-nucleotide polymorphism.

Acknowledgements

We thank Carol Ware, Angel Nelson, Jennifer Hesson and Chris Cavanaugh at the Institute for Stem Cell and Regenerative Medicine for providing us with the stem cells used in this study. This work was supported by grants from the National Institutes of Health (National Cancer Institute grant CA109597), the US Department of Defense (grant W81XWH-08-1-0636) and the John H. Tietze Foundation (to AK) and by a Mary Gates Endowment scholarship (to BJT).

Author details

[1]Institute for Stem Cell and Regenerative Medicine, University of Washington School of Medicine, 815 Mercer St., Seattle, WA 98109, USA. [2]Department of Genome Sciences, University of Washington, 3720 15th Ave NE, Seattle, WA 98195, USA. [3]National Centre for Biomedical Engineering Science, National University of Ireland, Galway, University Road, Galway, Republic of Ireland. [4]Department of Medicine (Endocrinology), Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA. [5]UW Medicine, South Lake Union, University of Washington School of Medicine, 815 Mercer St., Seattle, WA 98109, USA. [6]Division of Medical Genetics, Department of Medicine, University of Washington, 1705 NE Pacific St., Seattle, WA 98195, USA. [7]Department of Genetics, Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA. [8]Department of Medicine (Hematology), Albert Einstein College of Medicine, 1300 Morris Park Ave, Bronx, NY 10461, USA. [9]Department of Radiation Oncology, University of Washington, 1959 NE Pacific St., Seattle, WA 98195, USA.

Authors' contributions

AK conceived of and designed the study. BJT, EDR and AK performed the experiments. PÓB, AAG, JMG and NK provided bioinformatics support and carried out the statistical analyses. PW and KB contributed the samples. BJT, PW, AAG and AK drafted the paper. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 25 February 2011 Accepted: 3 August 2011
Published: 3 August 2011

References

1. Delaval K, Feil R: **Epigenetic regulation of mammalian genomic imprinting.** *Curr Opin Genet Dev* 2004, **14**:188-195.
2. Ferguson-Smith AC, Surani MA: **Imprinting and the epigenetic asymmetry between parental genomes.** *Science* 2001, **293**:1086-1089.
3. Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C: **High-resolution analysis of parent-of-origin allelic expression in the mouse brain.** *Science* 2010, **329**:643-648.
4. Chess A, Simon I, Cedar H, Axel R: **Allelic inactivation regulates olfactory receptor gene expression.** *Cell* 1994, **78**:823-834.
5. Bix M, Locksley RM: **Independent and epigenetic regulation of the interleukin-4 alleles in CD4$^+$ T cells.** *Science* 1998, **281**:1352-1354.
6. Holländer GA, Zuklys S, Morel C, Mizoguchi E, Mobisson K, Simpson S, Terhorst C, Wishart W, Golan DE, Bhan AK, Burakoff SJ: **Monoallelic expression of the interleukin-2 locus.** *Science* 1998, **279**:2118-2121.
7. Gimelbrant A, Hutchinson JN, Thompson BR, Chess A: **Widespread monoallelic expression on human autosomes.** *Science* 2007, **318**:1136-1140.
8. Reik W, Walter J: **Genomic imprinting: parental influence on the genome.** *Nat Rev Genet* 2001, **2**:21-32.
9. Bell AC, Felsenfeld G: **Methylation of a CTCF-dependent boundary controls imprinted expression of the *Igf2* gene.** *Nature* 2000, **405**:482-485.
10. Hark AT, Schoenherr CJ, Katz DJ, Ingram RS, Levorse JM, Tilghman SM: **CTCF mediates methylation-sensitive enhancer-blocking activity at the *H19/Igf2* locus.** *Nature* 2000, **405**:486-489.
11. Kanduri C, Pant V, Loukinov D, Pugacheva E, Qi CF, Wolffe A, Ohlsson R, Lobanenkov VV: **Functional association of CTCF with the insulator upstream of the *H19* gene is parent of origin-specific and methylation-sensitive.** *Curr Biol* 2000, **10**:853-856.
12. Phillips JE, Corces VG: **CTCF: master weaver of the genome.** *Cell* 2009, **137**:1194-1211.
13. Parelho V, Hadjur S, Spivakov M, Leleu M, Sauer S, Gregson HC, Jarmuz A, Canzonetta C, Webster Z, Nesterova T, Cobb BS, Yokomori K, Dillon N, Aragon L, Fisher AG, Merkenschlager M: **Cohesins functionally associate with CTCF on mammalian chromosome arms.** *Cell* 2008, **132**:422-433.
14. Rubio ED, Reiss DJ, Welcsh PL, Disteche CM, Filippova GN, Baliga NS, Aebersold R, Ranish JA, Krumm A: **CTCF physically links cohesin to chromatin.** *Proc Natl Acad Sci USA* 2008, **105**:8309-8314.
15. Stedman W, Kang H, Lin S, Kissil JL, Bartolomei MS, Lieberman PM: **Cohesins localize with CTCF at the KSHV latency control region and at cellular c-myc and *H19/Igf2* insulators.** *EMBO J* 2008, **27**:654-666.
16. Wendt KS, Yoshida K, Itoh T, Bando M, Koch B, Schirghuber E, Tsutsumi S, Nagae G, Ishihara K, Mishiro T, Yahata K, Imamoto F, Aburatani H, Nakao M, Imamoto N, Maeshima K, Shirahige K, Peters JM: **Cohesin mediates transcriptional insulation by CCCTC-binding factor.** *Nature* 2008, **451**:796-801.
17. Hadjur S, Williams LM, Ryan NK, Cobb BS, Sexton T, Fraser P, Fisher AG, Merkenschlager M: **Cohesins form chromosomal *cis*-interactions at the developmentally regulated *IFNG* locus.** *Nature* 2009, **460**:410-413.
18. Hou C, Dale R, Dean A: **Cell type specificity of chromatin organization mediated by CTCF and cohesin.** *Proc Natl Acad Sci USA* 2010, **107**:3651-3656.
19. Nativio R, Wendt KS, Ito Y, Huddleston JE, Uribe-Lewis S, Woodfine K, Krueger C, Reik W, Peters JM, Murrell A: **Cohesin is required for higher-order chromatin conformation at the imprinted *IGF2-H19* locus.** *PLoS Genet* 2009, **5**:e1000739.
20. Kacem S, Feil R: **Chromatin mechanisms in genomic imprinting.** *Mamm Genome* 2009, **20**:544-556.
21. Wen B, Wu H, Bjornsson H, Green RD, Irizarry R, Feinberg AP: **Overlapping euchromatin/heterochromatin-associated marks are enriched in imprinted gene regions and predict allele-specific modification.** *Genome Res* 2008, **18**:1806-1813.
22. Diss TC, Pan L, Peng H, Wotherspoon AC, Isaacson PG: **Sources of DNA for detecting B cell monoclonality using PCR.** *J Clin Pathol* 1994, **47**:493-496.
23. Nielsen J, Christiansen J, Lykke-Andersen J, Johnsen AH, Wewer UM, Nielsen FC: **A family of insulin-like growth factor II mRNA-binding proteins represses translation in late development.** *Mol Cell Biol* 1999, **19**:1262-1270.

24. Noubissi FK, Elcheva I, Bhatia N, Shakoori A, Ougolkov A, Liu J, Minamoto T, Ross J, Fuchs SY, Spiegelman VS: **CRD-BP mediates stabilization of *βTrCP1* and c-*myc* mRNA in response to β-catenin signalling.** *Nature* 2006, **441**:898-901.
25. Runge S, Nielsen FC, Nielsen J, Lykke-Andersen J, Wewer UM, Christiansen J: **H19 RNA binds four molecules of insulin-like growth factor II mRNA-binding protein.** *J Biol Chem* 2000, **275**:29562-29569.
26. Engel N, Thorvaldsen JL, Bartolomei MS: **CTCF binding sites promote transcription initiation and prevent DNA methylation on the maternal allele at the imprinted *H19/Igf2* locus.** *Hum Mol Genet* 2006, **15**:2945-2954.
27. Mahony S, Hendrix D, Golden A, Smith TJ, Rokhsar DS: **Transcription factor binding site identification using the self-organizing map.** *Bioinformatics* 2005, **21**:1807-1814.
28. Mahony S, Benos PV: **STAMP: a web tool for exploring DNA-binding motif similarities.** *Nucleic Acids Res* 2007, , **35 Web server:** W253-W258.
29. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B: **JASPAR: an open-access database for eukaryotic transcription factor binding profiles.** *Nucleic Acids Res* 2004, , **32 Database:** D91-D94.
30. Kim TH, Abdullaev ZK, Smith AD, Ching KA, Loukinov DI, Green RD, Zhang MQ, Lobanenkov VV, Ren B: **Analysis of the vertebrate insulator protein CTCF-binding sites in the human genome.** *Cell* 2007, **128**:1231-1245.
31. Zhang ZD, Rozowsky J, Snyder M, Chang J, Gerstein M: **Modeling ChIP sequencing in silico with applications.** *PLoS Comput Biol* 2008, **4**:e1000158.
32. Gombert WM, Krumm A: **Targeted deletion of multiple CTCF-binding elements in the human C-MYC gene reveals a requirement for CTCF in C-MYC expression.** *PLoS One* 2009, **4**:e6109.
33. Bock C, Reither S, Mikeska T, Paulsen M, Walter J, Lengauer T: **BiQ Analyzer: visualization and quality control for DNA methylation data from bisulfite sequencing.** *Bioinformatics* 2005, **21**:4067-4068.
34. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells.** *Cell* 2007, **130**:77-88.
35. Krumm A, Hickey LB, Groudine M: **Promoter-proximal pausing of RNA polymerase II defines a general rate-limiting step after transcription initiation.** *Genes Dev* 1995, **9**:559-572.
36. O'Brien T, Lis JT: **RNA polymerase II pauses at the 5' end of the transcriptionally induced *Drosophila hsp70* gene.** *Mol Cell Biol* 1991, **11**:5285-5290.
37. Zeitlinger J, Stark A, Kellis M, Hong JW, Nechaev S, Adelman K, Levine M, Young RA: **RNA polymerase stalling at developmental control genes in the *Drosophila melanogaster* embryo.** *Nat Genet* 2007, **39**:1512-1516.
38. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, Harp LF, Ye Z, Lee LK, Stuart RK, Ching CW, Ching KA, Antosiewicz-Bourget JE, Liu H, Zhang X, Green RD, Lobanenkov VV, Stewart R, Thomson JA, Crawford GE, Kellis M, Ren B: **Histone modifications at human enhancers reflect global cell-type-specific gene expression.** *Nature* 2009, **459**:108-112.
39. Mikkelsen TS, Xu Z, Zhang X, Wang L, Gimble JM, Lander ES, Rosen ED: **Comparative epigenomic analysis of murine and human adipogenesis.** *Cell* 2010, **143**:156-169.
40. Renda M, Baglivo I, Burgess-Beusse B, Esposito S, Fattorusso R, Felsenfeld G, Pedone PV: **Critical DNA binding interactions of the insulator protein CTCF: a small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci.** *J Biol Chem* 2007, **282**:33336-33345.
41. Kadota M, Yang HH, Hu N, Wang C, Hu Y, Taylor PR, Buetow KH, Lee MP: **Allele-specific chromatin immunoprecipitation studies show genetic influence on chromatin state in human genome.** *PLoS Genet* 2007, **3**:e81.
42. Kerkel K, Spadola A, Yuan E, Kosek J, Jiang L, Hod E, Li K, Murty VV, Schupf N, Vilain E, Morris M, Haghighi F, Tycko B: **Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation.** *Nat Genet* 2008, **40**:904-908.
43. Knight JC, Keating BJ, Rockett KA, Kwiatkowski DP: ***In vivo* characterization of regulatory polymorphisms by allele-specific quantification of RNA polymerase loading.** *Nat Genet* 2003, **33**:469-475.
44. Maynard ND, Chen J, Stuart RK, Fan JB, Ren B: **Genome-wide mapping of allele-specific protein-DNA interactions in human cells.** *Nat Methods* 2008, **5**:307-309.
45. McCann JA, Muro EM, Palmer C, Palidwor G, Porter CJ, Andrade-Navarro MA, Rudnicki MA: **ChIP on SNP-chip for genome-wide analysis of human histone H4 hyperacetylation.** *BMC Genomics* 2007, **8**:322.
46. Delaval K, Govin J, Cerqueira F, Rousseaux S, Khochbin S, Feil R: **Differential histone modifications mark mouse imprinting control regions during spermatogenesis.** *EMBO J* 2007, **26**:720-729.
47. Fournier C, Goto Y, Ballestar E, Delaval K, Hever AM, Esteller M, Feil R: **Allele-specific histone lysine methylation marks regulatory regions at imprinted mouse genes.** *EMBO J* 2002, **21**:6560-6570.
48. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells.** *Nature* 2007, **448**:553-560.
49. Ioannidis P, Kottaridi C, Dimitriadis E, Courtis N, Mahaira L, Talieri M, Giannopoulos A, Iliadis K, Papaioannou D, Nasioulas G, Trangas T: **Expression of the RNA-binding protein CRD-BP in brain and non-small cell lung tumors.** *Cancer Lett* 2004, **209**:245-250.
50. Ioannidis P, Mahaira L, Papadopoulou A, Teixeira MR, Heim S, Andersen JA, Evangelou E, Dafni U, Pandis N, Trangas T: **CRD-BP: a c-Myc mRNA stabilizing protein with an oncofetal pattern of expression.** *Anticancer Res* 2003, **23**:2179-2183.
51. Ioannidis P, Mahaira L, Papadopoulou A, Teixeira MR, Heim S, Andersen JA, Evangelou E, Dafni U, Pandis N, Trangas T: **8q24 copy number gains and expression of the c-myc mRNA stabilizing protein *CRD-BP* in primary breast carcinomas.** *Int J Cancer* 2003, **104**:54-59.
52. Ioannidis P, Trangas T, Dimitriadis E, Samiotaki M, Kyriazoglou I, Tsiapalis CM, Kittas C, Agnantis N, Nielsen FC, Nielsen J, Christiansen J, Pandis N: **C-*MYC* and IGF-II mRNA-binding protein (CRD-BP/IMP-1) in benign and malignant mesenchymal tumors.** *Int J Cancer* 2001, **94**:480-484.
53. Bieda M, Xu X, Singer MA, Green R, Farnham PJ: **Unbiased location analysis of E2F1-binding sites suggests a widespread role for E2F1 in the human genome.** *Genome Res* 2006, **16**:595-605.
54. Gombert WM, Farris SD, Rubio ED, Morey-Rosler KM, Schubach WH, Krumm A: **The c-*myc* insulator element and matrix attachment regions define the c-*myc* chromosomal domain.** *Mol Cell Biol* 2003, **23**:9338-9348.
55. Flanagin S, Nelson JD, Castner DG, Denisenko O, Bomsztyk K: **Microplate-based chromatin immunoprecipitation method, Matrix ChIP: a platform to study signaling of complex genomic events.** *Nucleic Acids Res* 2008, **36**:e17.
56. Dignam JD, Lebovitz RM, Roeder RG: **Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei.** *Nucleic Acids Res* 1983, **11**:1475-1489.
57. Chung JH, Bell AC, Felsenfeld G: **Characterization of the chicken β-globin insulator.** *Proc Natl Acad Sci USA* 1997, **94**:575-580.

doi:10.1186/1756-8935-4-14
Cite this article as: Thomas *et al.*: Allele-specific transcriptional elongation regulates monoallelic expression of the *IGF2BP1* gene. *Epigenetics & Chromatin* 2011 **4**:14.